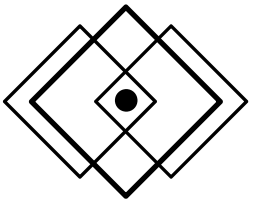
CHEMOENZYMATIC PEPTIDE SYNTHESIS THROUGH ENZYME-SPECIFIC ACTIVATION



ROSERI J. A. C. ROELOFSEN – DE BEER

The work presented in this thesis was conducted in the Synthetic Organic Chemistry group, Institute for Molecules and Materials at the Radboud University Nijmegen, Nijmegen, the Netherlands.

ISBN: 978-90-6464-571-6
Printed by: GVO drukkers & vormgevers B.V. | Ponsen & Looijen

NWO
Nederlandse Organisatie voor Wetenschappelijk Onderzoek

The research described in this thesis was financially supported by the programme Integration of Biosynthesis and Organic Synthesis (IBOS) of ACTS-NWO, and DSM (Geleen, the Netherlands).

Chemoenzymatic peptide synthesis through enzyme-specific activation

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S. C. J. J. Kortmann,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 5 september 2012
om 15.30 uur precies

door

Roseri Johanna Antonia Christina de Beer
geboren 6 juni 1983
te Tilburg

Promotor:	Prof. dr. F. P. J. T. Rutjes
Copromotoren:	Dr. S. B. Nabuurs Dr. F. L. van Delft
Manuscriptcommissie:	Prof. dr. J. C. M. van Hest Dr. M. C. R. Franssen (WUR) Prof. dr. D. B. Janssen (RuG)

Paranimfen:	Marieke Schor Michael J. P. M. de Beer

Chapter 6 Alcalase-catalysed peptide bond formation

Chapter 7 Synthesis

List of abbreviations

[α]	specific rotation
°C	degrees Celcius (centigrade)
Ac	acetyl
AcOH	acetic acid
Act	activating moiety
ACTS	Advanced Chemical Technologies for Sustainability
Ala	alanine (A)
aq.	aqueous
Ar	aryl
Arg	arginine (R)
Asn	asparagine (N)
Asp	aspartate (D)
Boc	*tert*-butoxycarbonyl
br (NMR)	broad
Bu	butyl
calcd	calculated
Cbz	benzyloxycarbonyl
CH_2Cl_2	dichloromethane
CLEA	crosslinked enzyme aggregate
compd	compound
CSPS	classical solution phase peptide synthesis
Cys	cysteine (C)
d	days
d (NMR)	doublet
DCC	N,N-dicyclohexylcarbodiimide
dd (NMR)	doublet of doublets
DIPEA	diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	*N,N*-dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dt (NMR)	doublet of triplets
DTT	dithiothreitol
e.g.	*exempli gratia* (for example)
EDC	*N*-ethyl-*N*′-(3-dimethyl aminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
equiv	equivalents
ESI	electrospray ionisation
Et	ethyl
et al.	*et alia* (and others)
EtOAc	ethyl acetate
Fmoc	9-fluorenylmethoxycarbonyl
GC	gas chromatography
Gln	glutamine (Q)
Glu	glutamate (E)
Gly	glycine (G)
h	hours
HEPES	*N*-(2-hydroxyethyl)piperazine-*N*'-ethanesulfonic acid
His	histidine (H)
HOBt	*N*-hydroxybenzotriazole
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
i.e.	*id est* (that is)
IBOS	Integration of Biosynthesis and Organic Synthesis
Ile	isoleucine (I)
IR	infrared
J (NMR)	coupling constant
LC-MS	liquid chromatography – mass spectrometry
Leu	leucine (L)
Lys	lysine (K)
M	molar
m	meta
m (NMR)	multiplet
m/z	mass to charge ratio
MD	molecular dynamics
Me	methyl
MeCN	acetonitrile
Met	methionine (M)
min	minutes
mp	melting point

mRNA	messenger RNA
MS	mass spectrometry
MTBE	methyl tert-butyl ether
$NaBH_3CN$	sodium cyanoborohydride
NGp	p-guanidinobenzyl amide
NMR	nuclear magnetic resonance
NPRS	nonribosomal peptide synthesis
Nu	nucleophile
NWO	Netherlands Organisation for Scientific Research
o	ortho
O3Cam	carbamoylpropyl ester
O3G	3-guanidinopropyl ester
O4A	4-aminobutyl ester
O4Cam	carbamoylbutyl ester
O4G	4-guanidinobutyl ester
O5A	5-aminopentyl ester
O5G	5-guanidinopentyl ester
OAb	p-amidinobenzyl ester
OBn	benzyl ester
OCam	carbamoylmethyl ester
ODmap	*p*-(dimethylamino)phenyl ester
ODmape	*p*-(dimethylamino)phenethyl ester
OGb	*p*-guanidinobenzyl ester
OGp	*p*-guanidinophenyl ester
OMe	methyl ester
Onb	nitrobenzyl ester
OTfe	2,2,2-trifluoroethyl ester
OTmap	*p*-(trimethylammonium) phenyl ester
OTmape	*p*-(trimethylammonium) phenethyl ester
p	para
$P_{1\text{-}4}$	position in peptide *N*-terminal of scissile bond
$P_{1\text{-}4}'$	position in peptide *C*-terminal of scissile bond
Ph	phenyl
Phe	phenylalanine (F)
*p*NA	p-nitroanilide
$POCl_3$	phosphoryl chloride
ppm	parts per million
Pro	proline (P)
*p*TSA	p-toluene sulfonic acid
q (NMR)	quartet
quant	quantitatively
Ra-Ni	Raney nickel
R_f	retention factor
RNA	ribonucleic acid
rt	room temperature
R_t	retention time
s (NMR)	singlet
S/H	synthesis over hydrolysis
$S_{1\text{-}4}$	subsite in enzyme *N*-terminal of scissile bond
$S_{1\text{-}4}'$	subsite in enzyme *C*-terminal of scissile bond
SAR	structure activity relationship
sat.	saturated
Ser	serine (S)
SPPS	solid phase peptide synthesis
t (NMR)	triplet
TBDMS	tert-butyldimethylsilyl
TBDPS	tert-butyldiphenylsilyl
^{t}Bu	tert-butyl
^{t}BuOH	tert-butanol
TEA	triethylamine
Tf	trifluoromethanesulfonyl
TFA	trifluoroacetic acid
THF	tetrahydrofuran
Thr	threonine (T)
TLC	thin layer chromatography
TMS	trimethylsilyl
tRNA	transfer RNA
Trp	tryptophane (W)
Trt	trityl (triphenylmethyl)
TS	transition state
Tyr	tyrosine (Y)
UV	ultraviolet
Val	valine (V)
X_{AA}	any amino acid

Chapter 1

Chemoenzymatic peptide synthesis

Numerous important processes in living organisms are mediated and regulated by peptides, which renders them interesting targets for the pharmaceutical, nutritional and cosmetic industry. As a result, exploring new ways to synthesise peptides is actively pursued. In this chapter both chemical and enzymatic peptide synthesis methods are summarised, with their respective advantages and disadvantages. There is a special focus on the substrate mimetic strategy: a chemoenzymatic peptide synthesis approach that potentially combines the best of two worlds. The definition of the subject and an outline of this thesis are included at the end of the chapter.

1.1 Introduction

Peptides are composed of amino acids that are linked to each other by amide bonds. The formation of these bonds is accompanied by the loss of water (Figure 1.1). Natural peptides consist of L-configured amino acids. When the rather arbitrary length of 50 amino acids is exceeded, the compound is usually referred to as a protein instead of a peptide. Amino acids can be abbreviated by a one- or three-letter code. Using the one-letter code, it is common to indicate the natural amino acids by a capital letter and the D-amino acids by a lower case letter. By convention, the amino terminus is taken as the beginning of a peptide chain and the carboxylic acid terminus as the end. The variation in size, shape, charge, hydrogen bonding capacity, hydrophobic character and chemical reactivity of amino acids accounts for the remarkably wide range of functions that are mediated by peptides. For example, peptides play a role in cell signalling and immune responses.

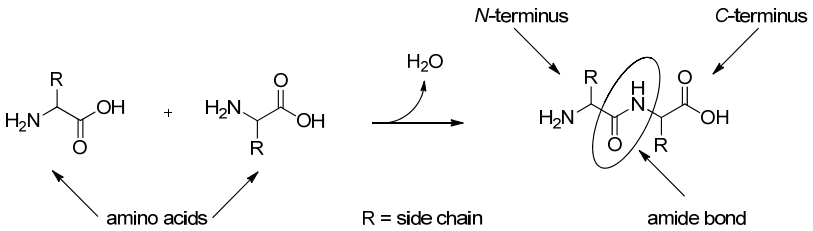


Figure 1.1 *Formation of a dipeptide from two amino acids.*

The fact that peptides influence many important processes in living organisms renders them interesting targets for pharmaceutical companies. However, untill recently they could not be applied because of their limited bioavailability. Since these problems are currently being addressed, peptides can really be developed and applied as therapeutics.[1] Other relevant applications of peptides include the use as nutrients[2] and cosmetic ingredients.[3] These recent developments have caused a market growth of (pharmaceutical) peptides, which in turn raised the challenge to synthetically take care of the increased demand, both in economic and environmental sense.

1.2 Large-scale peptide synthesis

Due to the fact that every method that exists to manufacture peptides on industrial scale has limitations and restrictions, the technologies for peptide production are not competitive with each other in most cases, but complementary. Prior to elaborating on each method (§ 1.3 and 1.4), an overview is provided of aspects that need to be considered, when choosing for a certain method (Table 1.1).

Table 1.1 *Methods and corresponding characteristics in industrial peptide synthesis*

	SPPS	CSPS	chemoenzymatic	fermentation
typical scale	mg - ton	g - ton	g - ton	g - ton
peptide length	med - large	short - med	short - large	short - med
sequence limitation	none	none	some	high
protection required	total	partial - total	minimal - partial	none
racemisation	some	some	none	none
development phase	mature	mature	embryonic	pacing
process R&D effort	very small	small	large	very large
production costs	very high	high	med	med – high
environmental impact	very high	high	some	some
more info in section	1.3.1	1.3.2	1.4.3	1.4.1

Besides the characteristics of the peptide such as length and sequence, costs, required quantity and the time to deliver the product are important parameters, which will differ from product to product. Andersson *et al.* focussed on practical aspects of chemical methods for the large-scale production of peptides including purification and isolation issues of final bulk products and regulatory considerations in an excellent review.[4]

1.3 Chemical peptide synthesis

Amino acids contain at least two reactive functional groups and can therefore react with each other in various ways, resulting in an uncontrolled peptide product. To prevent this problem, protecting groups are applied. A distinction is made between 'temporary' protection of the α-amino function and 'permanent' protection of the side chain functionalities. In other words, the side-chain protection stays in place during the whole synthesis, while the *N*-terminal protection is removed after every coupling step to enable the attachment of a new amino acid. Protecting groups that are not influenced by mutual deprotection conditions are called orthogonal. The more complicated the peptide target molecule, the more sophisticated protection strategies are required and therefore the development of new protecting groups has been deeply tied to peptide synthesis. Significant advances have been made in this area and comprehensive analysis of amino acid protecting group strategies is provided in the review of Isidro-Llobet.[5]

The reaction between carboxylic acids and amines does not occur spontaneously at ambient temperature. The required elimination of water is generally established by means of a coupling reagent, which converts the acid into an activated acylating species with a good leaving group, prior to treatment with the amine. Over the past decades, studies to identify the most efficient and highest yielding coupling reagent, with lowest epimerisation levels and byproduct formation, have resulted in an overwhelming choice. Both Valeur *et al.* and El Faham *et al.* reviewed the scope and limitations of recently developed coupling reagents.[6]

1.3.1 Solid phase peptide synthesis (SPPS)

The most commonly used strategy for peptide synthesis is SPPS, which was applied for the first time by Merrifield.[7] He attached a growing peptide chain with its *C*-terminus to a completely insoluble but swellable polymeric support and alternated the deprotection of the *N*-terminus of the peptide with the coupling of a new protected amino acid (Figure 1.2). Extensive washing removed the reagents between each successive coupling. A colouring test[8] was applied to check the completeness of each coupling.

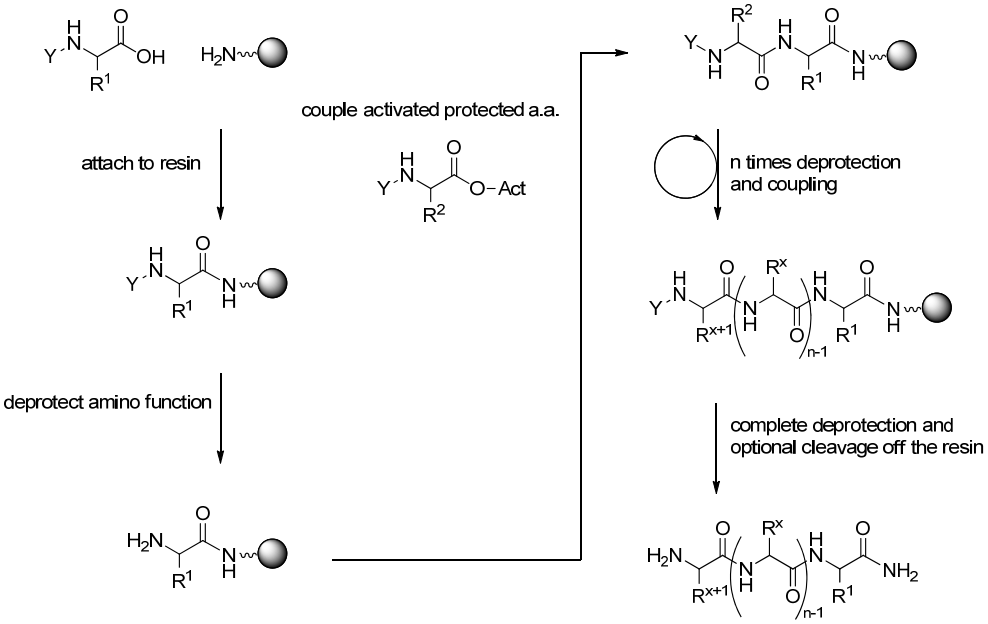


Figure 1.2 *Principle of Solid Phase Peptide Synthesis. The temporary protecting group Y is selectively removed, exposing a new reactive site, which is then coupled to the next protected amino acid; R^1, R^2, R^x, $R^{x=1}$= amino acid side chains, if necessary, protected with permanent protecting group.*

The original Merrifield strategy is based on gradual acid lability and utilises the Boc protecting group for temporary protection of the α-amino function. Permanent side-chain protection usually involves benzylic or related groups that are cleaved with strong acids. The concept was however extended with the development of new protecting groups. In

Fmoc chemistry, the base labile Fmoc group protects the α-amino function.[9] The permanent side-chain protection usually relies on the *t*-Bu group or other acid labile groups. Currently, the Fmoc strategy is generally applied, because it involves milder reagents and often leads to peptides of higher purity and in higher yields.

Besides the synthesis strategy, the nature of the solid carrier is a relevant parameter in SPPS. Solid support particles should preferably be of conventional and uniform size, mechanically robust, easily filterable, chemically inert and stable under the conditions of synthesis. They must also be accessible to the solvents allowing the penetration of reagents and the enlargement of the peptide chain within its microstructure. Furthermore, the solid phase should not physically interact with the growing peptide chain.[10]

SPPS has important advantages over synthesis in solution (§ 1.3.2), and represented such a step forward that Merrifield was awarded the Nobel Prize in chemistry in 1984. These advantages include:

- Reactions can be driven to completion through the use of excess reagents.
- Excess reagents and soluble by-products can be simply removed by resin washing.
- Physical losses are minimal since the product remains attached to the polymer throughout the synthesis.

This type of peptide synthesis has one major disadvantage, being the production of more than stoichiometric amounts of environmentally aggressive and toxic waste.[10-11] Another drawback of the SPPS strategy entails the inability to purify intermediate products; only after the final cleavage step purification is possible.

1.3.2 Classical solution phase peptide synthesis (CSPS)

CSPS was the first general method developed for peptide synthesis. Small peptides, consisting of only a few amino acids, were built in solution, usually in EtOAc or DMF. Minimal side-chain protection strategies can be applied, because fully protected amino acids or peptides are highly insoluble, resulting in various undesired side reactions. Yet, the main advantage of this technique is that the intermediate products can be isolated and purified after each step of synthesis.

Besides stepwise elongation, CSPS is nowadays frequently applied to couple fragments, often obtained by SPPS. During the synthesis of long or highly polar peptide sequences by SPPS, the yield drops drastically after approximately 30 residues. A convergent approach, in which multiple peptide segments are condensed in a hierarchical fashion to form the full-length polypeptide chain, is generally more efficient than a sequential approach.[12] Fragment condensation suffers from two major drawbacks: to avoid racemisation of the *C*-terminal amino acid, only glycine or proline can be used in this position. Furthermore, the fragment with the free *N*-terminal amine has to be used in excess compared to the other fragment, to afford a high-yielding reaction.

1.3.3 Native chemical ligation

The previously described methods could not completely satisfy the need to obtain sufficient peptide or protein material with the potential to site-selectively introduce chemical functionalities for chemical biology research. This desire has been the driving force for the development of various chemoselective ligation and modification strategies,[13] of which native chemical ligation is by far the most advanced technique.[14]

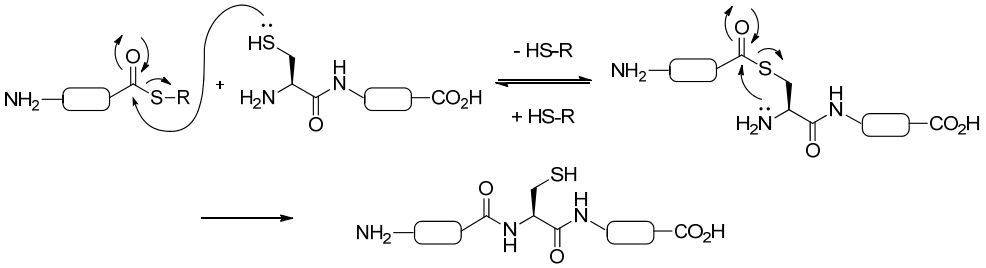


Figure 1.3 *The two-step mechanism of native chemical ligation*

In native chemical ligation a native peptide bond is formed in aqueous solution at neutral pH from unprotected peptides, which are easy to handle because of their high solubility. One fragment is C-activated as a thioester, while the other fragment contains an *N*-terminal cysteine (Figure 1.3). These groups selectively react with each other in a two-step fashion, thereby preserving the integrity of the unprotected side chains. The first step involves a reversible thiol-thioester exchange, followed by rapid intramolecular S→N acyl transfer, via a favourable five-membered transition state. The reaction rate can be increased by the addition of an excess of a more reactive arylthiol. Importantly, additional internal Cys residues in the peptide do not interfere with the overall reaction pathway, since the rearrangement requires a terminal nucleophilic amine instead of an internal amide. All twenty amino acids are allowed at the *C*-terminus, albeit that the ligation rate is significantly affected by the different side chains.

The high stability and accessibility of the starting materials is key for the success of this strategy. The main initial limitation, being the presence of a relatively rare cysteine in the sequence, can be overcome by desulfurisation to an alanine residue or the use of an auxiliary strategy. A complete overview of the scope and limitations in chemical ligation is provided by Kent.[14]

1.4 Enzymatic peptide synthesis

The current chemical methods generate large amounts of waste because often an excess of reagents has to be used to drive the reaction to completion and a copious volume of organic solvent is part of the process. An enzymatic approach would offer an attractive alternative, since enzymes preferably act in water, require no or minimal side-chain protection, and make the use of reactive chemicals superfluous. Obviously, enzymatic reactions are associated with drawbacks on a different level, which need to be overcome.

1.4.1 Recombinant expression

Protein synthesis in living organimsms is an enormously complex process, involving numerous components such as an mRNA template, tRNA, amino acyl transferases and ribosomes. It is therefore difficult and expensive to use the enzymes involved outside their natural context in a reaction flask. It is however possible to use microorganisms for this purpose, since they contain the complete synthesis apparatus. By genetic modification, the DNA encoding the peptide of interest is introduced into the host cell, for example the *Escherichia coli* bacterium. Subsequent transcription, translation and expression leads to the peptide of interest, which can then be isolated. This technique is particularly suitable for large natural peptides and proteins. Disadvantages of this method are the restriction that only natural amino acids can be incorporated and the high costs for development and production.

1.4.2 Nonribosomal peptide synthesis (NRPS)

Nonribosomal peptides represent a class of bioactive peptides, displaying antimicrobial, antiviral, immunosuppressive, and antitumor activities, containing unique structural features such as heterocyclic elements, D-amino acids, and glycosylated as well as *N*-methylated residues. They are not produced by ribosomal synthesis, but large multi-enzyme complexes, which simultaneously represent template and biosynthetic machinery, are responsible for their assembly. These complexes are organised in a modular fashion and for each specific building block a specific module exists, consisting of catalytically independent domains that regulate substrate recognition, activation, binding, modification, elongation and release. Due to the increasing knowledge and current techniques, it is now possible to rationally redesign the NRPS complex to synthesise a particular peptide target. A comprehensive review about NRPS and its potentials has been produced by Sieber.[15]

1.4.3 Isolated proteases

Besides enzymes dedicated to peptide synthesis, also proteases, which naturally hydrolyse peptide bonds, can be employed for the same goal under carefully tuned circumstances. They are convenient because proteases (1) are active at mild conditions, with pH optima in the range of 6 to 8, (2) are robust and stable, (3) do not require stoichiometric cofactors and (4) are also highly stereo- and regioselective.

1.4.3.1 Classifications and mechanisms[16]

Based on their catalytic mechanism, proteolytic enzymes are divided in six families: serine, cysteine and threonine proteases on the one hand, and aspartate, glutamate[17] and metalloproteases on the other hand. In the first category, the amino acids after which the families are named represent the nucleophile in the active site. As a typical example, the catalytic triad of a serine protease is depicted in Figure 1.4. Histidine and aspartate shuttle away the proton of the serine residue by hydrogen-bonding interactions, rendering it more nucleophilic and capable of attacking the peptide substrate. The negative charge that

develops during the formation of the tetrahedral complex is stabilised by the oxyanion hole. Upon the collapse of this transition state, part of the peptide is cleaved off and a covalent acyl-enzyme intermediate is formed. Subsequent nucleophilic attack by water liberates the product and regenerates the enzyme.

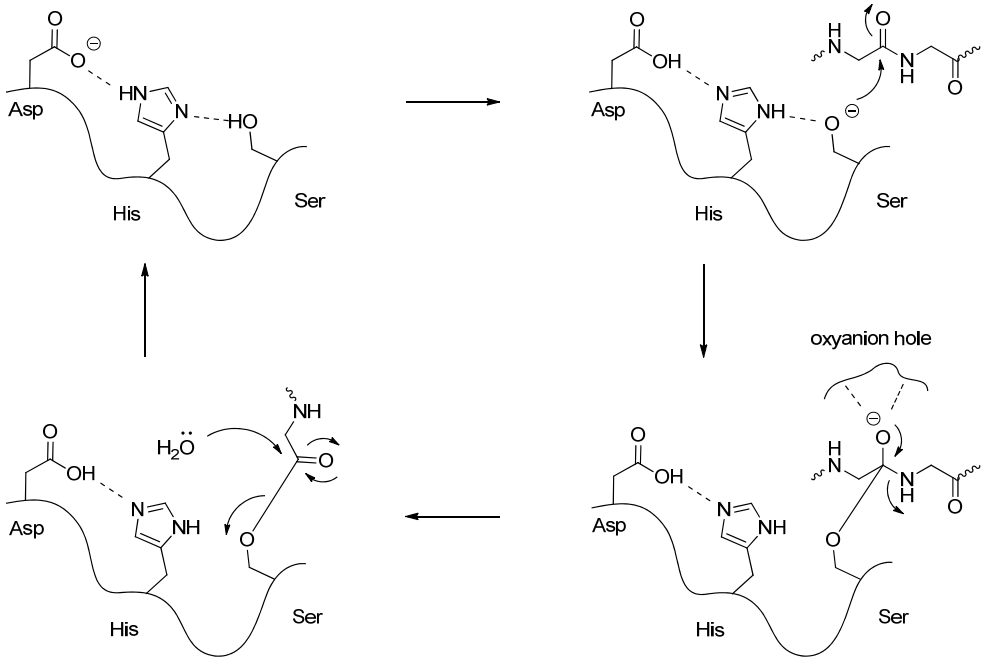


Figure 1.4 *The catalytic mechanism of serine proteases*

In the second category, an activated water molecule serves as the nucleophile. In metalloproteases activation comes from coordination to a metal ion, mostly zinc, sometimes also cobalt (Figure 1.5). In addition, the metal coordinates to the oxygen of the carbonyl in the substrate, giving rise to a more electrophilic carbon. Basic or acidic amino acid residues from the enzyme fill the remaining available coordination sites.[18]

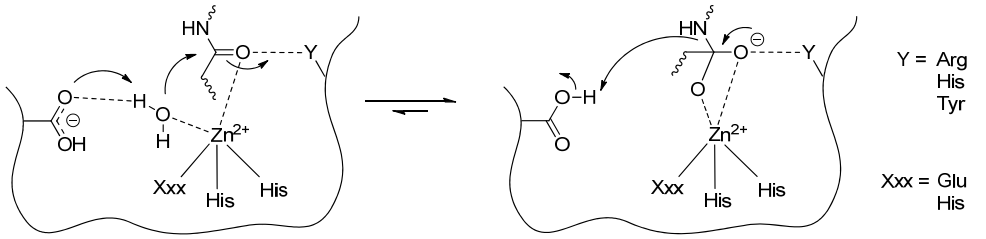


Figure 1.5 *The catalytic mechanism of metalloproteases*

As indicated by the names, in aspartic or glutamic acid proteases the respective amino acids are involved in catalysis. The commonly accepted mechanism involves general acid-

base catalysis, which is aided by another aspartate residue in the first case and a glutamine in the latter (Figure 1.6). The carboxylate activates the water by abstracting a proton generating a hydroxide. Electrophilic assistance by polarising the carbonyl bond of the scissile peptide is provided by the second catalytic amino acid.

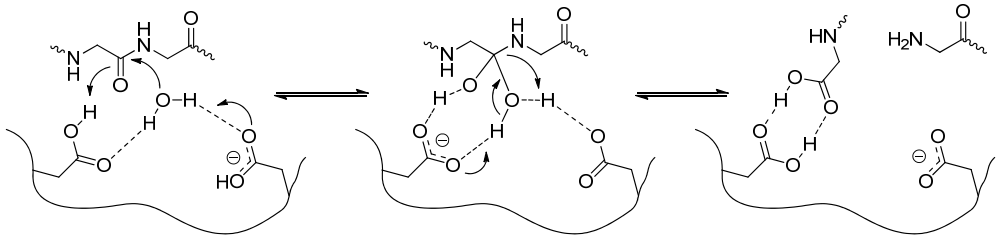


Figure 1.6 *The catalytic mechanism of aspartic proteases*

Another way of categorisation is based on the cleavage site in the protein. Endopeptidases hydrolyse amide bonds within a peptide or protein, as opposed to exopeptidases that act directly near the termini. Aminopeptidases release a single amino acid from the *N*-terminus, whereas carboxypeptidases liberate a single amino acid from the *C*-terminus. Amidases cleave terminal amides, releasing a carboxylate and ammonia.

The exact site of hydrolysis is determined by the specificity of the protease against amino acid residues on either side of the scissile bond. Schechter and Berger[19] introduced comprehensive nomenclature (Figure 1.7) to describe the interactions between the peptide (P) and the enzyme subsites (S) taking this splitting point as the starting point for counting.

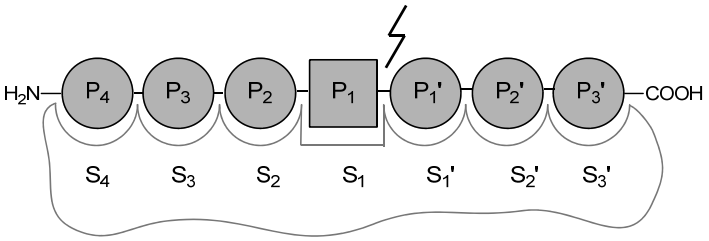


Figure 1.7 *Schechter and Berger nomenclature for the active site of a protease and the peptide substrate. The scissile bond is located between P_1 and P_1'.*

Characteristics concerning classification, mechanism and specificity of individual proteases (papain, trypsin, α-chymotrypsin and alcalase) will be introduced in the corresponding chapters.

1.4.3.2 Thermodynamic vs. kinetic control

When proteases are used under physiological conditions, they perform hydrolysis and the reversed synthesis reaction is negligible. Since proteases, like all catalysts, only alter the rate with which the thermodynamic equilibrium of the reaction is reached and do not change the equilibrium itself, the reaction conditions need to be manipulated to shift the

equilibrium in favour of the synthesis reaction. In practice two synthetically useful strategies are applied *i.e.* the thermodynamically and the kinetically controlled approach.

In the thermodynamically controlled synthesis of peptides, the product is obtained by the direct reverse of the hydrolysis reaction. According to the principle of microscopic reversibility, both the formation and the hydrolysis of the peptide bond proceed by the same mechanism and through the same intermediate (Figure 1.8). This implies that acyl donors with a free carboxylate function are used, and all proteases can be applied, independent of their individual catalytic mechanism. The rate limiting step is represented by the formation of the acyl intermediate from the carboxylic acid.

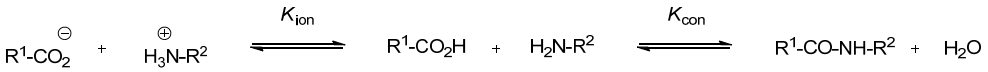


Figure 1.8 *Thermodynamically controlled synthesis*

In general, the thermodynamic approach gives low yields and requires high enzyme concentrations. The unfavourably positioned equilibrium can be shifted in the direction of synthesis by the addition of organic solvents, which influences the equilibrium constants of both ionisation (K_{ion}) and conversion (K_{con}). If the dielectric constant of the medium is lowered, the acidity of the carboxyl group of the acyl donor is also reduced, resulting in the presence of more protonated, uncharged species. Furthermore, the water activity is diminished by the addition of cosolvents, and, as a result, the equilibrium is shifted towards synthesis. Conducting the reaction in the absence of water, *i.e.* in a dry aprotic solvent, would be the ultimate situation in this respect. A similar increase of K_{con} can be achieved when the product is removed from the reaction mixture by precipitation or a two-layer system where the product can diffuse to the organic layer, while the enzymatic reaction takes place in aqueous environment. A major downside of these manipulations is the probable instability and inactivity of the enzyme.

The main issue in kinetically controlled peptide synthesis is the difference in reaction rate of the various possible reaction partners. Slightly activated acyl donors (Ac-X), such as esters, are normally used. First, a tetrahedral enzyme-substrate complex (E···Ac-X) is formed, that collapses into the covalent acyl-enzyme intermediate (Ac-E) (Figure 1.9), thus limiting this approach to serine and cysteine proteases. The deacylation of this intermediate can be performed by all the nucleophiles (H_2O or Nu) present in the reaction mixture. Water and the nucleophilic amine will compete with each other for the transfer of the acyl group, resulting in hydrolysis (Ac-OH) or synthesis (Ac-Nu), respectively. The nucleophilic attack is the rate limiting step.

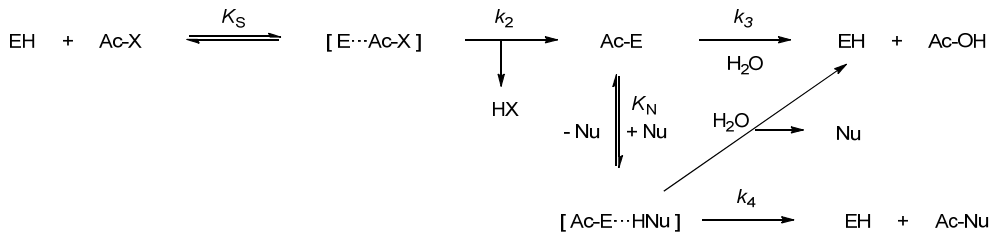


Figure 1.9 *Kinetically controlled synthesis*

Compared to thermodynamic control, the kinetic approach shows relatively high reaction rates and requires a minimal amount of enzyme. The yield of peptide product will depend on the apparent ratio of transferase-to-hydrolase rate constants. A drawback of kinetic control is the necessity to terminate the reaction after the acyl donor is consumed, because the synthesis product will be hydrolysed again.

There are several ways to affect kinetically controlled reactions, although obviously the limits allowed by the enzyme need to be taken into account. The pH can be raised, since the amine is only nucleophilic in unprotonated form. Increasing the substrate concentrations or the enzyme-to-substrate ratio will also be beneficial.

1.4.4 Improved properties of biocatalysts

Instability and inactivity of enzymes caused by, for example, the addition of organic solvents or by a high operating pH can be dealt with in various ways. Two of them will be discussed, *i.e.* immobilisation and enzyme engineering. Both approaches have the additional advantage that they can also influence the S/H ratio of a protease.

1.4.4.1 Immobilisation

Different types of immobilisation can be distinguished, including the involvement of a solid support, the entrapment or encapsulation using polymers and the direct crosslinking of enzymes with a bifunctional reagent.[20] A further subdivision for solid carriers is based on the type of interaction with the enzyme, which can be non-covalent adsorption by hydrophilic, hydrophobic or stronger ionic interactions, or the formation of covalent linkages. In what way an enzyme exactly binds, is mainly random, although more techniques are developed that are able to direct it. The most distinctive aspect of entrapment or encapsulation is the protection it offers by preventing direct contact with the environment. With crosslinking the problem of reduced specific and volumetric activity of the biocatalyst is circumvented. A practical example is the CLEA method – crosslinked enzyme aggregates– in which precipitated enzyme is connected with glutaraldehyde. Whatever the nature of immobilisation, an important requirement for efficient biocatalysis is that substrates and products can freely diffuse inside and outside the immobilisation support. Which of the previously mentioned techniques will give the best results, is hard to predict. Hanefeld *et al.* suggest factors that need to be taken into

account when planning the immobilisation of an enzyme and provide some guidelines for the selection of a suitable technique.[20a]

The increased stability of enzymes by immobilisation is, in all probability, caused by the restricted freedom of movement and the directly related decreased chance of unfolding. The enzyme is trapped in a certain conformation, which likely also is at the basis of the changed substrate specificity that is sometimes observed upon immobilisation. An added advantage of immobilised catalysts is that they can easily be separated from the reaction mixture, which facilitates product purification and is additionally convenient when the enzyme must be reused.

1.4.4.2 Enzyme engineering

Biocatalyst engineering by a genetic approach is a powerful tool to create catalysts with tailored properties, such as high stability, but also improved specificity, substrate scope, enantioselectivity and diminished product inhibition. Reetz described the most recent developments in this field in an excellent review, and, although the focus is on improving the enantioselectivity of enzymes, the underlying strategies and added protocols are informative and generally applicable.[21]

In the strategy of directed evolution, natural evolution is imitated by subjecting the gene encoding the enzyme of interest to repeated cycles of gene mutagenesis. Both the error-prone polymerase chain reaction[22] and DNA shuffling[23] can be used for this purpose. The generated gene library is then inserted in a bacterial host and expressed. In a screening procedure the candidates with the desired features are selected and the process is continued until the desired degree of improvement has been achieved. An overwhelming amount of enzyme variants can be created this way, requiring an efficient and effective screening method.

The input of knowledge in the mutagenesis process can structurally reduce the number of variants that need to be produced and screened. This approach is called site directed mutagenesis. Based on structural, mechanistic or bioinformatics data, putative randomisation sites are selected, which can reasonably be expected to be involved in a certain property. Often these sites will be located around the active site, but distal residues that may induce allosteric effects are viable options as well. In iterative site directed mutagenesis the putatively important sites are mutated in a systematic way, in order to take cooperative effects into account.

1.5 Substrate mimetics approach

1.5.1 Principle and mechanism

The basis for the substrate mimetics approach was laid in 1977 by establishing that inverse ester substrates containing a cationic centre in the ester leaving group instead of in the amino acid side chain, can be hydrolyzed by trypsin.[24] Schellenberger was the first to apply this concept in a reversed enzymatic synthesis reaction. Only in 1997, the term ‘substrate mimetics’ was introduced by Bordusa.[25]

According to literature, the substrate mimetics strategy combines the advantages of both enzymatic and chemical peptide synthesis, while most of the disadvantages are solved. The advantages include freedom in the choice of amino acid to be coupled, high stereo- and regioselectivity and the need for only small amounts of reagents and solvents. The latter two benefits can be realised by using enzymes for the synthesis reaction. The broad recognition of many amino acids, with a single enzyme, can be acquired with a trick. Normally, a protease recognises the side chain of a specific amino acid, and cleaves the substrate at the *C*-terminal site of this residue. When this recognition element is, however, mimicked as the site-specific ester leaving group, the substrate is still accepted by the enzyme, making recognition independent of the side chain of the amino acid.[26] A well-known example is the guanidinophenyl (OGp) group, a mimic of the arginine side chain, which is naturally recognised by the enzyme trypsin. The catalytic mechanism[25] of this substrate mimetic (SM; Figure 1.10 B) is analogous to the natural situation (Figure 1.10 A).

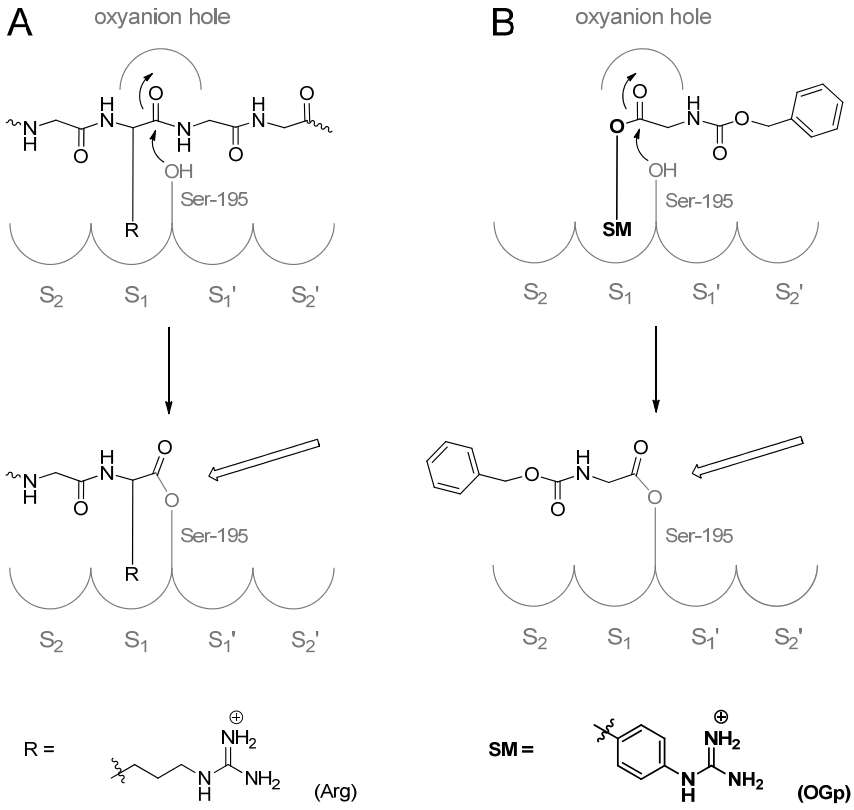


Figure 1.10 *Catalytic mechanism of the serine protease trypsin with (A) a natural substrate and (B) a substrate mimetic. Recognition takes place in the S_1 subsite, with subsequent nucleophilic attack of Ser-195 onto the carbonyl. The substrate mimetic is bound in a reversed orientation compared to the orientation of the natural substrate. The covalent intermediate can only be liberated by nucleophilic attack from the S' region (double arrow).*

First the substrate mimetic binds to the S_1 pocket, which is followed by nucleophilic attack of the catalytic serine residue on the carbonyl of the substrate. Kinetic and computational studies have indicated that the substrate mimetic should bind in a reverse orientation to enable this attack.[27] In this way, the developing negative charge can still be stabilised by the oxyanion hole. Because the deacylation step requires unoccupied S' subsites, the non-specific acyl residue has to flip from the S' to the S subsites prior to nucleophilic attack, in order to liberate both product and enzyme. Water as the nucleophile will result in the hydrolysis product, while an amino acid nucleophile will lead to dipeptide synthesis.

1.5.2 Known enzyme – mimetic combinations

In addition to OGp, several variations of this moiety have been investigated to probe the structural requirements of a mimetic for trypsin (Figure 1.11).[24, 28]

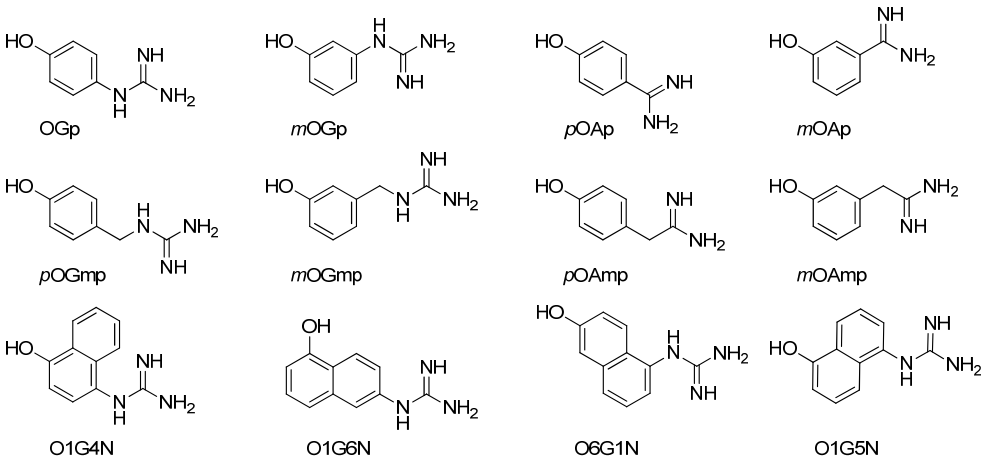


Figure 1.11 *Structures of efficient substrate mimetics and less efficient variations*

Besides OGp, *p*OAp was found to be efficient in enzymatic peptide synthesis. The relocation of the guanidino or amidino group to the meta-position (*m*OGp and *m*OAp) resulted in much less effective esters. The altered bond angle may be an important determinant here, as well as the changed direction between the carbonyl group and the cationic substituent. The introduction of a methylene between the cationic substituent and the aromatic ring (OGmp and OAmp) was highly unfavourable, probably due to the different leaving group ability of the phenol as a result of the non-conjugated amidino or guanidino group. From the naphthalene derivatives, only O1G4N, in which the two substituents are aligned with the shorter axis of the ring, gave a good peptide coupling yield. O1G5N appeared to be moderately efficient, whereas both O1G6N and O6G1N reacted very slowly.

Furthermore, the applicability of the OGp ester was established for the cysteine protease clostripain, which specifically recognises arginine residues.[29] This enzyme possesses wide acyl acceptor tolerance, since branched amines, amino alcohols and cyclic amines were all accepted.

A fruitful attempt to extend the substrate mimetic strategy to α-chymotrypsin was accompanied by computational docking studies.[30] It was shown that OGp can also be employed for this enzyme –which is specific for large aromatic residues– in spite of the presence of a charged guanidine group. Rationalisation by a docking approach provided a model that allows potential mimetics to be predicted. The binding energies of empirically chosen leaving groups derived from benzyl (OBn, OPic), indolyl (OInd) or phenyl (OPh, OCp and OCap) esters (Figure 1.12) were calculated and successfully correlated to the observed specificity constants in a hydrolysis experiment. The OCap ester of Boc-L-Ala-OH was identified as an efficient and slightly higher specific substrate mimetic for α-chymotrypsin than the OGp ester.

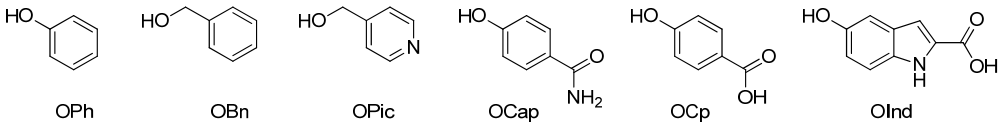


Figure 1.12 *Structures of potential mimetics used in a computational approach*

To demonstrate that the substrate mimetics strategy is not restricted to OGp or arginine-recognising proteases, novel mimetics were designed for two enzymes, V8 protease from *Staphylococcus aureus* and Glu-specific endoprotease from *Bacillus licheniformis*, with a strong preference for glutamate in their recognition pocket (Figure 1.13).[31]

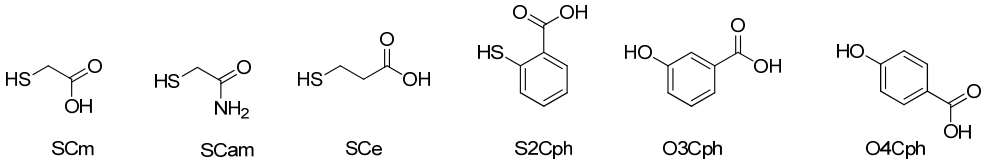


Figure 1.13 *Structures of potential mimetics for Glu-specific proteases*

Both biocatalysts displayed a similar preference pattern for these compounds, but the latter enzyme appeared to react more efficiently. The carboxymethyl thioester (SCm) was effective in acyl transfer reactions, independent of the acyl moiety connected to it. However, compared to classical substrates, the enzyme requirement was high. The negative charge was essential for recognition as was demonstrated by the complete inactivity of the uncharged carboxyamide (SCam). In an effort to optimise the mimetic, the alkyl chain was extended by one carbon (SCe) to make it more similar to a glutamate. In addition, the alkyl chain was replaced by a phenyl linker (Cph), analogous to the situation that OGp is a good arginine mimetic, the position of the carboxylate was changed (2Cph, 3Cph, 4Cph) or the thioester bond was replaced by an ester bond (OCph). None of these changes significantly affected the properties of the mimetic.

1.5.3 Limitations and possible solutions

Although the substrate mimetics strategy has many advantages, some intrinsic drawbacks connected to enzymatic peptide synthesis still remain. The primary specificity of a protease continues to be present with the result that product hydrolysis can occur when the specific amino acid is part of the peptide, which is most likely the case for long peptides. Moreover, the specificity for the acyl donor and the preference for a nucleophile are individual characteristics for each enzyme, therefore the efficiency of S/H and thus the potential for peptide synthesis will greatly differ from one protease to another. In recent years much effort has been devoted to overcome these limitations.

1.5.3.1 Biocatalyst engineering

As mentioned in § 1.4.4.2, an effective way to change the properties of an enzyme is by engineering, either chemically or genetically. Here, examples with respect to substrate mimetics are discussed. A chemical method comprises the methylation of the N^{ε} active site histidine in trypsin, thereby generating an enzyme variant in which the amidase activity is lost, but the esterase activity is retained.[32] Peptide formation with this catalyst using OGp esters was successful and the products were not subject to secondary hydrolysis by the enzyme. Another way to obtain a universally applicable trypsin variant with decreased cleavage activity is site-directed mutagenesis.[33] The known trypsin mutant D189S was taken as a starting point, because it exhibits diminished native amidase activity for trypsin substrates, whereas its proteolytic activity towards chymotrypsin substrates is improved by one order of magnitude. The total hydrolysing capacity of the enzyme was further repressed based on a thorough computational analysis of the binding site. With the double mutations D189A and S190A, both involved in determining substrate specificity in trypsin, a synthetically useful catalyst was acquired for substrate mimetics-mediated peptide ligations. In order to find an enzyme with a specificity tailored at the artificial substrate mimetic moiety OGp, a variation on the site-directed mutagenesis strategy was applied.[34] The rationally designed D189E mutation in trypsin was anticipated to disturb the interaction of the carboxylate with arginine or lysine by the presence of an additional methyl group, while not hindering the interaction with OGp too much. This manipulation indeed led to increased specificity towards OGp esters compared to arginyl- or lysyl-bonds. Yet a different approach is the search for enzymes with improved properties from different sources. Amongst others, various bacterial and mammalian trypsins have been investigated, as well as trypsin enzymes from cold-adapted fish, as they display substantially higher catalytic efficiency than their mammalian counterparts.[28c, 35] An elegant idea to circumvent the whole problem and to take advantage of the specificity is to synthesise all-D-peptides with an L-specific enzyme or apply this concept the other way around.[36]

1.5.3.2 Freezing strategy or ionic liquids

Two other approaches to suppress the undesired enzymatic hydrolysis reaction of specific peptide bonds have been described, both of which involve lowering the availability of water. By freezing the reaction mixture, the amidase activity of a protease for classical substrates is severely restricted, while the inherent esterase activity is much less affected. Based on experiments with trypsin, chymotrypsin and *Bacillus* lichenisformis Glu-specific endopeptidase and the appropriate substrate mimetics, it was concluded that the effects for substrate mimetics are even more pronounced. Peptides containing specific amino acids are generally hydrolysed with higher preference than the respective substrate mimetics. At –15 °C, however, such specific fragments could be synthesised without the expected proteolytic side reaction, albeit in a much slower rate. Furthermore, the ratio between hydrolysis and aminolysis of the substrate mimetics was improved.[37]

The application of ionic liquids as reaction medium additive in substrate mimetic-mediated peptide synthesis was positively evaluated. Trypsin, chymotrypsin and V8 protease and the appropriate substrate mimetics were investigated. Wild-type biocatalysts could be used, since any competitive proteolytic side-reactions were suppressed. Furthermore, compared to the addition of classical organic solvents, high turnover rates were obtained. Due to the exceptional properties of ionic liquids chemically labile reactants remained highly stable; moreover the solubility of many reactants is vastly improved in this medium.[38]

1.6 Computational modelling[39]

1.6.1 Molecular docking

The interaction between an enzyme and its substrate is often a black box for a synthetic chemist. Preparing a range of substrate analogues and experimentally determining their activity should hopefully result in a relevant structure activity relationship (SAR). A computational tool that can help to overcome the lack of insight into the interaction is molecular docking. In molecular docking, knowledge about the enzyme is used to predict and optimise the binding of small molecules. The primary requirement is a high resolution model of the enzyme structure. When the position of the active site in the structure is known, it is called direct docking. As in enzymatic models, where more flexibility was introduced by switching from the traditional 'lock and key' concept to the induced fit model, flexibility has been increasingly implemented in molecular docking software. Nowadays, both the substrate and to a growing extent also the enzyme are regarded as flexible. The main restriction here is the limited computational feasibility. The number of degrees of freedom that have to be considered grows exponentially with the number of accessible enzyme conformations.[40]

1.6.2 MD simulations

More insight into the time dependent behaviour of a molecular system can be obtained by molecular dynamics (MD) simulations, which provide detailed information on the

fluctuations and conformational changes of proteins or protein-ligand complexes. Molecular mechanic principles lie at the basis of the calculations, meaning that Newton's equation of motion is applied: **F=m•a**, where **F** is the force exerted on the particle, **m** is its mass and **a** is its acceleration. The initial velocities on the atoms are randomly generated at the beginning of the dynamics run and force fields are used to describe the interactions between the atoms involved. These force fields are an empirical substitution for the potential energy function with parameters that are optimised on the basis of reality, with consideration of explicit or implicit solvent whatever is desired. Newton's equation is iteratively solved at regular time intervals, resulting in a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. The cycle is repeated for a predefined number of steps. Parameterisation of the force field plays a crucial role in the accuracy of the molecular model predictions. Furthermore, it is important to sample a sufficient amount of phase space and acquire a representative conformational sampling.

1.6.3 *Ab initio* calculations

When the mechanism of enzyme catalysis is the subject of study, *ab initio* calculations should be employed, because this technique is suitable for describing electronic processes such as charge transfer, transition states and reaction paths.[41] It is based on quantum mechanical calculations instead of empirical parameters. In the Hartree-Fock method, a molecular system is regarded as a linear combination of atomic orbitals, which, in turn, can be described by a basis set of wave functions. With the density functional theory the properties of a many-electron system can be determined by using functionals of the spatially dependent electron density. In both cases, the computational requirement is large, therefore this method is only feasible for relatively small systems, containing around maximally hundred atoms. Therefore quantum mechanical calculations are mostly restricted to the active site of an enzyme, while the complementary use of highly efficient force field based molecular mechanics methods for the surroundings can provide the overall picture.

1.7 Problem definition

1.7.1 Context provided by IBOS Programme

IBOS (Integration of Biosynthesis & Organic Synthesis) is a research programme under the auspices of NWO (Netherlands Organisation for Scientific Research), in particular the division ACTS-CW. The latter focuses on societally relevant chemical issues in which the Dutch government, knowledge institutions and the Dutch chemical and life sciences industry are involved. A change of strategy in industrial synthetic chemistry by integrating state of the art organic chemistry and modern biochemistry and biotechnology is the aim of IBOS. This will enable a new and sustainable future for industrial synthesis of complex products, such as pharmaceuticals, nutrients or new materials, which contribute considerably to the quality of life. In addition, a drastic reduction of waste is envisioned.

One of the themes within the IBOS programme is chemoenzymatic peptide synthesis. Peptides form the most promising class of bioactive compounds for future biomedical and pharmaceutical applications. However, purely chemical synthesis suffers from serious disadvantages, such as low yields and high process costs. It is therefore of great value to implement solutions offered by nature in this process, as most of the biochemical processes that occur in living organisms easily outperform the 'best practices' of conventional synthetic methods. The aim of the project is to develop low-cost versatile chemoenzymatic routes for the synthesis of natural and non-natural peptides of importance for pharmaceutical and/or food applications. Key objectives are:

- Discovery and engineering of peptidases and amidases that can catalyse *C*-terminal activation and coupling of all common amino acids;
- Defining and optimising suitable bioreactor configurations and process conditions ;
- Development and validation of the industrial applicability of peptide synthesis based on substrate mimetic methods.

These topics were distributed between the collaboration partners of this project. At the University of Groningen under the supervision of Prof. dr. D.B. Janssen, Ana Toplak has been working on the cloning and expression of new and better endoproteases using two approaches. The first approach focused on subtiligase as a starting point, and the second on obtaining new subtilisin-like enzymes from extremophiles. The production of a peptide amidase and the subsequent characterisation and determination of the biocatalytic potential for industrial application has been carried out by Irfan Arif. At Wageningen University under the supervision of Prof. dr. ir. J. Tramper, Petra Vossenberg is engaged in the development of process conditions for optimal peptide coupling, preferably in one pot, investigating the influence of water activity and enzyme incompatibility. The third topic, exploration of the substrate mimetics strategy, is the subject of this thesis and the objectives and chosen approach are detailed in the next section.
Industrial partner DSM Innovative Synthesis B.V., represented by dr. P. J. L. M. Quaedflieg, contributed 'in-house' and 'in-kind' research efforts. Part of this research was executed by PhD-student Timo Nuijens, who focussed on peptide synthesis in organic media using activated esters and immobilised Alcalase.

1.7.2 Objective and approach

The principle of the substrate mimetics approach seems promising according to literature, but since it was only demonstrated on a small scale with expensive mimetic groups, the viability for industrial scale needs to be probed. To that end, the scope and limitations of the substrate mimetics strategy to enzymatically couple amino acids in aqueous media was investigated, paying particular attention to the following aspects:

- the use of cheap enzymes with a broad substrate range
- the design of simple mimetics, which are synthetically readily accessible and preferably can be synthesised enzymatically

In addition, it would be of great value to know which requirements a substrate mimetic has to meet in order to function well and to understand their mode of action. Several computational modelling techniques will be applied to gain insight into the active site and obtain valuable input for the design of new mimetics.

1.7.3 Outline of thesis

Chapter 1 serves as an introduction into chemoenzymatic peptide synthesis, providing an overview of the advantages, potentials, drawbacks and limitations of chemical versus enzymatic peptide synthesis. Special attention is given to the substrate mimetics strategy. With the intention to widen its scope, this strategy was applied to the cysteine protease papain, a cheap enzyme with broad specificity. These results complemented with computational studies are described in Chapter 2. In Chapter 3, simplified esters are developed, based on the observation that in combination with papain, enzyme-specific activation is a more appropriate term than substrate mimetics. Chapter 4 focusses on determining what properties are essential for developing a good activating ester or mimetic using trypsin as the protease. The influence of different enzymes on enzyme-specific activation is studied comparing α-chymotrypsin and papain, which is described in Chapter 5. In Chapter 6, the effect of nearly anhydrous reaction conditions on enzyme-specific activation is considered using Alcalase-CLEA. Chapter 7 describes the synthesis of all compounds that were required for the previous chapters. By way of overall conclusion, the results are put in perspective in Chapter 8.

1.8 References

[1] (a) R. E. W. Hancock, H.-G. Sahl, *Nat. Biotechnol.* **2006**, *24*, 1551-1557; (b) D. Hans, P. R. Young, D. P. Fairlie, *Med. Chem.* **2006**, *2*, 627-646; (c) P. Vlieghe, V. Lisowski, J. Martinez, M. Khrestchatisky, *Drug Discovery Today* **2010**, *15*, 40-56; (d) A. A. Zompra, A. S. Galanis, O. Werbitzky, F. Albericio, *Future Med. Chem.* **2009**, *1*, 361-377.

[2] (a) R. Hartmann, H. Meisel, *Curr. Opin. Biotechnol.* **2007**, *18*; (b) V. A. Tutel'yan, V. K. Khavinson, V. V. Malinin, *Bull. Exp. Biol. Med.* **2003**, *135*.

[3] (a) M. P. Lupo, A. L. Cole, *Dermatol. Ther.* **2007**, *20*, 343–349; (b) B. S. Oricha, *Afr. J. Pharm. Pharmacol.* **2010**, *4*, 127-129.

[4] L. Andersson, L. Blomberg, M. Flegel, L. Lepsa, L. Nilsson, M. Verlander, *Biopolymers (Pept. Sci.)* **2000**, *55*, 227–250.

[5] A. Isidro-Llobet, M. Álvarez, F. Albericio, *Chem. Rev.* **2009**, *109*, 2455–2504.

[6] (a) E. Valeur, M. Bradley, *Chem. Soc. Rev.* **2009**, *38*, 606–631; (b) A. El-Faham, F. Albericio, *Chem. Rev.* **2011**, *111*, 6557–6602.

[7] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.

[8] (a) W. S. Hancock, J. E. Battersby, *Anal. Biochem.* **1976**, *71*, 260-264; (b) E. Kaiser, R. L. Colescott, C. D. Bossinger, P. I. Cook, *Anal. Biochem.* **1970**, *34*, 595-598; (c) T. Vojkovsky, *Pept. Res.* **1995**, *8*, 236-237.

[9] L. A. Carpino, G. Y. Han, *J. Am. Chem. Soc.* **1970**, *92*, 5748-5749.

[10] F. Guzmán, S. Barberis, A. Illanes, *J. Biotechnol.* **2007**, *10*, 279-314.

[11] D. Kumar, T. C. Bhalla, *Applied Microbiology & Biotechnology* **2005**, *68*, 726-736.

[12] J. Y. Lee, D. Bang, *Biopolymers (Pept. Sci.)* **2010**, *94*, 441-447.

[13] C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem. Int. Ed.* **2008**, *47*, 10030-10074.

[14] S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338-351.

[15] S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **2005**, *105*, 715-738.

[16] A. J. Barret, N. D. Rawlings, J. F. Woessner, *Handbook of Proteolytic Enzymes*, Elsevier, London, **2004**.

[17] M. Fujinaga, M. M. Cherney, H. Oyama, K. Oda, M. N. G. James, *Proc. Natl. Acad. Sci. USA.* **2004**, *101*, 3364-3369.

[18] M. Hernick, C. A. Fierke, *Arch. Biochem. Biophys.* **2005**, *433*, 71-84.

[19] I. Schechter, A. Berger, *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157-162.

[20] (a) U. Hanefeld, L. Gardossi, E. Magner, *Chem. Soc. Rev.* **2009**, *38*, 453–468; (b) D. Brady, J. Jordaan, *Biotechnol. Lett.* **2009**, *31*, 1639-1650.

[21] M. T. Reetz, *Angew. Chem. Int. Ed.* **2011**, *50*, 138-174.

[22] (a) D. W. Leung, E. Chen, D. V. Goeddel, *Technique* **1989**, *1*, 11-15; (b) C. Cadwell, G. F. Joyce, *PCR Methods Applic.* **1992**, *2*, 28-33.

[23] W. P. C. Stemmer, *Nature* **1994**, *370*, 389 – 391.

[24] K. Tanizawa, Y. Kasaba, Y. Kanaoka, *J. Am. Chem. Soc.* **1977**, *99*, 4485-4488.

[25] F. Bordusa, D. Ullmann, C. Elsner, H. D. Jakubke, *Angew. Chem. Int. Ed.* **1997**, *36*, 2473-2475.

[26] F. Bordusa, *Curr. Protein Pept. Sci.* **2002**, *3*, 159-180.

[27] M. Thormann, S. Thust, W. H. Hofmann, F. Bordusa, *Biochemistry* **1999**, *38*, 6056-6062.

[28] (a) K. Itoh, H. Sekizaki, E. Toyota, N. Fuijiwara, K. Tanizawa, *Bioorg. Chem.* **1996**, *24*, 59-68; (b) H. Sekizaki, K. Itoh, A. Shibuya, E. Toyota, K. Tanizawa, *Chem. Pharm. Bull.* **2007**, *55*, 1514-1517; (c) H. Sekizaki, K. Itoh, E. Toyota, K. Tanizawa, *Chem. Pharm. Bull.* **1998**, *46*, 846-849; (d) H. Sekizaki, K. Itoh, E. Toyota, K. Tanizawa, *Amino Acids* **1999**, *17*, 285-291.

[29] R. Günther, A. Stein, F. Bordusa, *J. Org. Chem.* **2000**, *65*, 1672-1679.

[30] (a) R. Günther, C. Elsner, S. Schmidt, H. J. Hoffman, F. Bordusa, *Org. Biomol. Chem.* **2004**, *2*, 1442-1446; (b) R. Günther, S. Thust, H. J. Hofmann, F. Bordusa, *Eur. J. Biochem.* **2000**, *267*, 3496-3501.

[31] (a) N. Wehofsky, F. Bordusa, *FEBS Lett.* **1999**, *443*, 220-224; (b) N. Wehofsky, J. D. Wissmann, M. Alisch, F. Bordusa, *Biochim. Biophys. Acta* **2000**, *1479*, 114-122.

[32] K. Itoh, H. Sekizaki, E. Toyota, K. Tanizawa, *Bioorg. Chem.* **1997**, *25*, 307-319.

[33] R. Grünberg, I. Domgall, R. Günther, K. Rall, H. J. Hofmann, F. Bordusa, *Eur. J. Biochem.* **2000**, *267*, 7024-7030.

[34] S. Xu, K. Rall, F. Bordusa, *J. Org. Chem.* **2001**, *66*, 1627-1632.

[35] (a) T. Fuchise, H. Kishimura, H. Sekizaki, Y. Nonami, G. Kanno, S. Klomklao, S. Benjakul, B.-S. Chun, *Food Chem.* **2009**, *116*, 611-616; bH. Sekizaki, M. Murakami, K. Itoh, E. Toyota, K. Tanizawa, *J. Mol. Catal. B: Enzym.* **2000**, *11*, 23-28; (c) E. Toyota, D. Iyaguchi, H. Sekizaki, M. Tateyama, K. K. S. Ng, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2009**, *65*, 717–723.

[36] (a) N. Wehofsky, S. Thust, J. Burmeister, S. Klussmann, F. Bordusa, *Angew. Chem. Int. Ed.* **2003**, *42*, 677-679; (b) N. Wehofsky, A. Pech, S. Liebscher, S. Schmidt, H. Komeda, Y. Asano, F. Bordusa, *Angew. Chem. Int. Ed.* **2008**, *47*, 5456 –5460.

[37] (a) N. Wehofsky, M. Haensler, S. W. Kirbach, W. J.D., F. Bordusa, *Tetrahedron: Asymmetry* **2000**, *11*, 2421-2428 ; (b) N. Wehofsky, S. W. Kirbach, M. Haensler, J. D. Wissmanne, F. Bordusa, *Org. Lett.* **2000**, *2*, 2027-2030.

[38] N. Wehofsky, C. Wespe, V. Cerovsky, A. Pech, E. Hoess, R. Rudolph, F. Bordusa, *ChemBioChem* **2008**, *9*, 1493-1499.

[39] H.-D. Höltje, W. Sippl, D. Rognan, G. Folkers, *Molecular modeling. Basic principles and applications. 3rd Edition*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, **2008**.

[40] S. B. Nabuurs, M. Wagener, J. de Vlieg, *J. Med. Chem.* **2007**, *50*, 6507–6518.

[41] R. Lonsdale, K. E. Ranaghan, A. J. Mulholland, *Chem. Commun.* **2010**, *46*, 2333-2512.

Chapter 2

Papain-catalysed peptide bond formation: enzyme-specific activation with guanidinophenyl esters

The substrate mimetics approach is a versatile method for small-scale enzymatic peptide-bond synthesis in aqueous systems. The protease-recognised amino acid side chain is incorporated in an ester leaving group, the substrate mimetic. This shift of the specific moiety enables the acceptance of amino acids and peptide sequences that are normally not recognised by the enzyme. The guanidinophenyl group (OGp), a known substrate mimetic for the serine proteases trypsin and chymotrypsin, has now been applied for the first time in combination with papain, a cheap and commercially available cysteine protease. To provide insight into the binding mode of various Z-X_{AA}-OGp esters, computational docking studies were performed. The results strongly point at enzyme-specific activation of the OGp esters in papain through a novel mode of action, rather than their functioning as mimetics. Furthermore, the scope of a model dipeptide synthesis reaction was investigated with respect to both the amino acid donor and the nucleophile. Molecular dynamics simulations were carried out to prioritise 22 natural and unnatural amino acid donors for synthesis. Experimental results correlate well with the predicted ranking and show that nearly all amino acids are accepted by papain.

This chapter has been published:
R. J. A. C. de Beer, B. Zarzycka, H. I. V. Amatdjais-Groenen, S. C. B. Jans, T. Nuijens, P. J. L. M. Quaedflieg, F. L. van Delft, S. B. Nabuurs, F. P. J. T. Rutjes, *ChemBioChem* **2011**, *12*, 2201-2207.

2.1 Introduction

The importance of peptides in the fields of healthcare and nutrition renders the amide bond probably the most synthesised chemical bond.[1] In light of this, both chemocatalytic and enzymatic strategies for amide-bond formation are being developed. The use of enzymes is advantageous, particularly on an industrial scale, because they are usually selective, prevent racemisation, and require minimal or no side-chain protection. The typical enzymes of choice for peptide synthesis are proteases, readily available enzymes that normally hydrolyse peptidic amide bonds, but also show the capacity to effect amide bond formation. The typical selection criterion for a specific protease is based on the specificity for a particular amino acid residue on either side of the scissile bond. However, the main drawback of such an approach is the need for a different enzyme for nearly each desired peptide bond, as is nicely illustrated by the enzymatic synthesis of octapeptide CCK-8 for which three different proteases were required.[2] Secondary hydrolysis of the reaction products necessitated a judicious choice in the order of the fragment couplings, and prudent fine-tuning of the reaction conditions for each individual step. These drawbacks in combination with the limited recognition of many amino acids, both natural and unnatural, severely restrict the universal application of enzymes.

In the so-called substrate mimetics strategy, the problem of limited applicability is overcome by incorporating the enzyme-recognised amino acid side chain moiety in the ester leaving group, thereby making the enzymatic recognition independent of the side chain. A well-known example is the guanidinophenyl (OGp) group, a mimic of the arginine side chain, which is naturally recognised by the protease trypsin.[3] Besides fundamental studies on the mechanism[4] (see § 1.5.1), much research on substrate mimetics has been devoted to method improvement and to solving difficulties related to enzymatic peptide synthesis. For example, several variations of the OGp ester have been evaluated as alternative substrate mimetics for trypsin,[5] the OGp ester was applied to different enzymes such as chymotrypsin[6] and clostripain,[7] but also to trypsin variants with diminished hydrolytic activity[8] or purified from different exotic sources.[5c, 5d, 9] The undesired enzymatic hydrolysis reaction of specific peptide bonds can be successfully suppressed by freezing the reaction mixture[10] or by using ionic liquids as solvents.[11] Furthermore, it has been shown that new substrate mimetics could be designed for two enzymes that recognise negatively charged amino acids,[12] thus indicating that more-general application of the strategy might be feasible. Despite the considerable amount of existing research, the substrate mimetics approach has never been applied to the cysteine protease papain. This cheap and commercially available enzyme would be a good candidate, also for industrial application, because it has been effectively used in enzymatic peptide synthesis before and cysteine proteases are known to result in a better synthesis-over-hydrolysis (S/H) ratio than serine proteases.[13] The slight preference of papain for arginine and lysine residues suggested that the existing OGp ester could be used as a mimetic. Furthermore, papain has a broader specificity than trypsin, and this could provide relevant information about the applicability of the substrate mimetics approach to

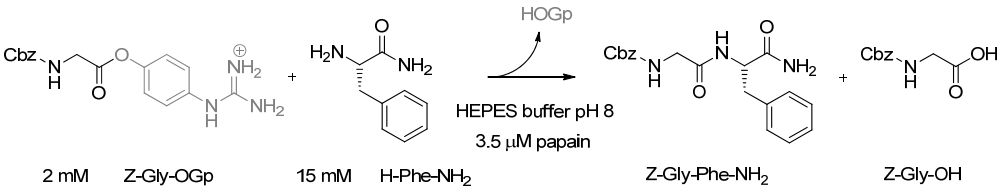
less specific enzymes. So far only proteases that bear a narrow specificity towards distinct amino acid side chains have been used; hence this research could potentially widen the scope of the substrate mimetics approach.

We used the OGp ester as a potential substrate mimetic for papain-induced dipeptide synthesis under aqueous conditions. Subsequent docking studies provided insight into the binding mode, and, based on these results, MD simulations served to predict a set of suitable amino acid donors. The resulting ranking was experimentally verified, and the scope of the amino acid acceptor was determined.

2.2 Results & Discussion

2.2.1 Initial dipeptide synthesis reaction

Z-Gly-OGp was prepared as a test substrate in order to study the OGp ester as a mimetic applied to papain. Z-Gly-OH was esterified with *p*-[*N*', *N*''-Di(Boc)guanidino]phenol[14] by using DCC as coupling reagent. Glycine was chosen to be incorporated in the acyl donor because it is the simplest amino acid and is not recognised by papain. Next, Z-Gly-OGp was subjected to the enzymatic reaction with papain in which H-Phe-NH_2 was used as the acyl acceptor because of its clear UV-visibility at 254 nm, which simplified the HPLC analyses (Scheme 2.1).



Scheme 2.1 *Enzymatic synthesis with papain*

To our delight, the dipeptide product was formed quickly (20 min) and in high yield (92%) with hardly any enzymatic hydrolysis (2.4%). This compares favourably to results obtained with trypsin.[8a] Concurrently, we observed that chemical hydrolysis of the starting material under exactly the same reaction conditions but in the absence of papain was substantial (5.6%). This spontaneous hydrolysis is clearly undesirable, because it will increase with longer reaction times.

2.2.2 Prediction of the binding mode and validation of Z-X_{AA}-OGp esters

Having shown that papain can recognise Z-Gly-OGp, we developed an increased molecular understanding of the substrate mimetics approach in papain. A molecular modelling study was performed by using the flexible docking programme Fleksy.[15] The results were visualised and analysed by using the YASARA programme.

Crystallisation and subsequent structure determination by X-ray methods revealed that papain is a single-chain polypeptide containing three disulfide bridges, which is folded into two domains with the active site in a groove between them.[16] The most important

residues in the active site are Cys25, which provides the nucleophilic thiol, and His159, which completes the catalytic diad (Figure 2.1 A). The oxyanion hole is formed by the side chain NH_2 group of Gln19 and the backbone amide of Cys25. Residues Asp158, Gly66 and Trp177 are involved in a conserved hydrogen-bonding network that is required for substrate affinity.[17]

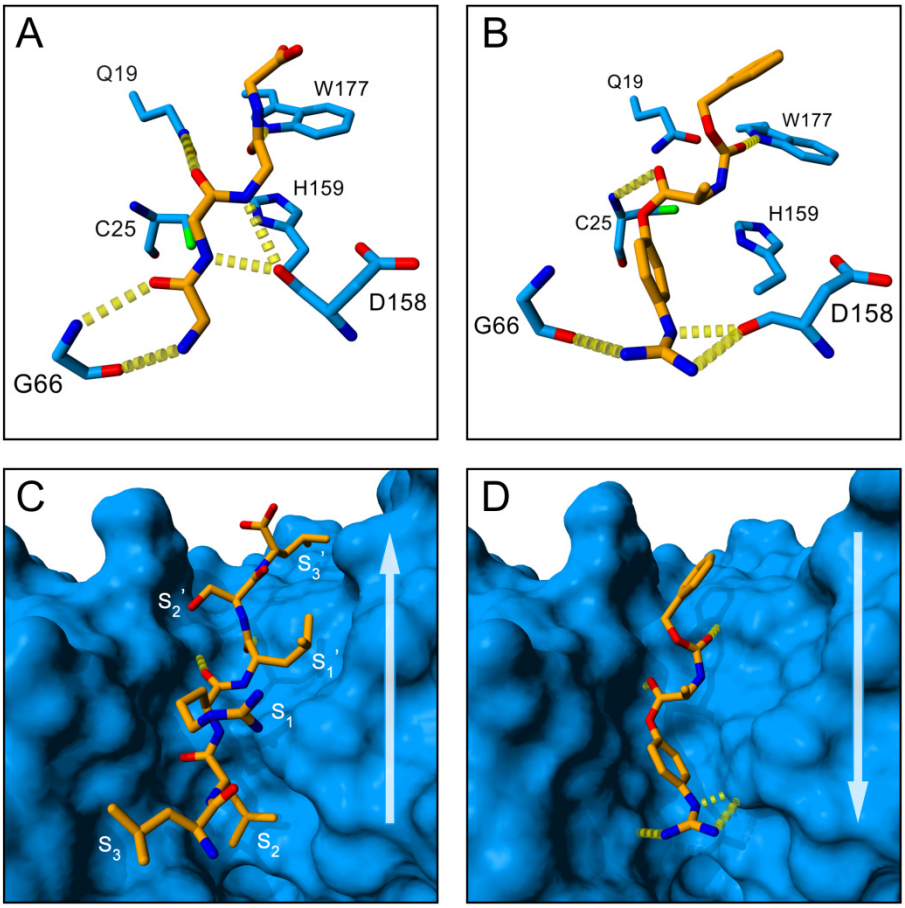


Figure 2.1 *Molecular modelling of papain substrates. Hydrogen bonding interactions to functionally important amino acids in the papain active site are shown for (A) the backbone of a peptide substrate and (B) the compound Z-L-Ala-OGp. The overall orientation in the papain peptide binding groove is shown for (C) the hexapeptide LLRLSL and (D) the compound Z-L-Ala-OGp. The different papain subsites occupied by the hexapeptide are indicated. The white arrows indicate the* N- *to* C-*terminal direction of the scissile bond, thus illustrating the 'reversed' binding mode of Z-L-Ala-OGp.*

To gain insight into the influence of the configuration of the amino acids and sterically demanding substrates, in addition to the achiral glycine ester, the OGp esters of L-alanine, D-alanine, β-alanine and L-proline were considered in the docking study. Figure 2.1 B shows that Z-L-Ala-OGp perfectly fits in the active site, interacting in a similar manner as

the natural substrate (Figure 2.1 A). Based on the obtained Z-L-Ala-OGp binding mode, we rationalised that the oppositely configured D-alanine ester would be a much poorer substrate, as its methyl side chain is predicted to clash with the catalytic His-159. This effect is even more profound and disadvantageous for the secondary amino acid L-proline. In case of the β-alanine ester, the interactions with the oxyanion hole and the hydrogen bonding network with the guanidino group could be maintained, but the beneficial interaction with Trp177 was disrupted by the introduction of an extra carbon atom in the ester backbone. Experimental verification of these modelling-based hypotheses confirmed that reaction rates for Z-D-Ala-OGp and Z-β-Ala-OGp are indeed around 100 times lower than those for Z-Gly-OGp and Z-L-Ala-OGp and that Z-Pro-OGp does not react at all. This all provides experimental support for the binding mode of Z-L-Ala-OGp proposed in Figure 2.1 B.

The different view in Figure 2.1 C shows that the natural substrate is positioned in the groove with all the amino acid side chains exposed, including the specificity-determining arginine. The substrate mimetic is also located in this cleft, but is oriented reversely as compared to the natural substrate (Figure 2.1 D), similar to what Bordusa described for the OGp ester in trypsin.[18] It is remarkable though, that the guanidino group present in the mimetic is situated in the groove, unlike the natural substrate's arginine side chain. This suggests that, in the case of papain, the OGp ester is not actually mimicking the natural substrate in the sense that the guanidino group binds at the same position. Nonetheless, the OGp ester is specifically recognised by the enzyme, and therefore we propose to call it an enzyme-specific activating ester.

2.2.3 Molecular dynamics simulations of Z-X_{AA}-OGp esters in papain

To further probe the viability of enzyme-specific activating esters as a more general enzymatic method for peptide synthesis, we investigated whether the successful dipeptide synthesis with Z-Gly-OGp and H-Phe-NH_2 could be extended to a variety of other natural and unnatural amino acids. Figure 2.1 D clearly shows that the side chain of Z-L-Ala-OGp points up and out of the groove, thereby implying there should be sufficient space to accommodate the side chains of other natural amino acids as well. Rather than determining the scope of the acyl donor directly, we incorporated a computational approach as an intermediate step. A reliable model would be of great use in predicting which amino acids will be accepted by the enzyme papain in order to prioritise them for synthesis and for reaction planning. All 20 natural and the two previously evaluated unnatural amino acids were docked, and the stabilities of the resulting complexes were assessed with the help of molecular dynamics simulations.

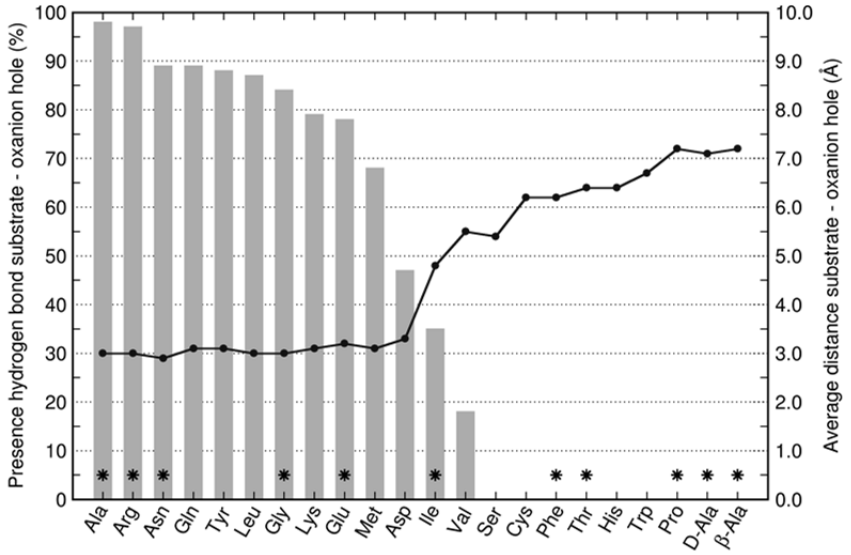


Figure 2.2 *Stability of modelled papain-Z-X_{AA}-OGp complexes as assessed by molecular dynamics simulations. The presence of a hydrogen bond between Gln19 of the papain oxyanion hole and carbonyl of the scissile bond of the substrate throughout the simulation time is shown by bars (left axis), and the average distance between the oxyanion hole (Gln19) and carbonyl of the scissile bond of the substrate throughout the simulation is shown by a line (right axis). The amino acids that were selected for synthesis and testing are indicated with asterisks. All amino acids were considered in the L-conformation, unless otherwise indicated.*

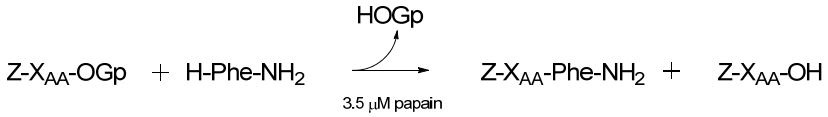
This was done on the assumption that the aforementioned interactions with the oxyanion hole are essential to the formation of a productive enzyme-substrate complex. Both the distance to oxyanion-hole residue Gln19 and the percentage of simulation time for which this residue forms a hydrogen bond with the substrate, were used as measures of complex stability. Initially oxyanion hole Cys25 was also taken into account, but including these interactions did not affect the results. The ranking is mostly determined by the specific interactions that the side chains make with papain during the simulation (Figure 2.2). As a result, amino acids with similar characteristics generally show a comparable ranking. For instance, asparagine and glutamine are positioned close to each other, as are the β-branched amino acids isoleucine and valine. The observed discrepancy between phenylalanine and tyrosine can be explained by the fact that, although the docking showed a nice fit, phenylalanine is drawn into the nearby hydrophobic S_1' pocket during the MD simulation, an event that is probably prevented for tyrosine due to the presence of the additional polar hydroxyl group. Based on the docking studies alone, one would expect a higher ranking for serine, but this is presumably prohibited by the formation of an intramolecular hydrogen bond during the simulation.

2.2.4 Verification of the MD-based predictions

To verify the MD-based rankings and to determine the scope of the acyl donor experimentally, a representative set of amino acids was selected for validation (marked with an asterisk in Figure 2.2). The corresponding Z-X_{AA}-OGp esters were synthesised starting from amino acids with appropriate side-chain protection. Yields of the coupling reactions varied between 63 and 98%. The acidic deprotection appeared to be troublesome in some cases, because chemical hydrolysis of the ester occurred as a side reaction (<10%). Nevertheless, the crude product obtained after Boc deprotection was used directly in the enzymatic reactions, and a correction was performed afterwards. The enzymatic reaction was monitored for three hours. The identity of the products formed in the enzymatic reaction was confirmed by chemical synthesis of reference compounds and LC-MS analysis. Table 1 presents either the time to reach full conversion of the OGp esters, or the conversion after three hours. The indicated percentages of enzymatic synthesis and hydrolysis remained constant over time, as measured after 24 hours unless stated otherwise. The S/H ratio for the various amino acids gives an ambiguous impression. While excellent in the case of glycine and threonine (entries 4 and 8), proline (entry 9) does not react with papain at all. With the OGp esters of both arginine and asparagine (entries 2 and 3) non-enzymatic cyclic side products were formed, a piperidone[19] and succinimide[20] respectively; this was facilitated by the guanidinophenyl leaving group. Over time the piperidone side product was converted into Z-Arg-OH, while the succinimide side product concentration remained constant.

It is not trivial to assess the correlation between the computational results in Figure 2.2 and the experimental results in Table 2.1. Given the large variation in the S/H ratios for the various amino acids, the analysis should be restricted to conversion rates only. This can be rationalised by realising that the simulations address only the effective formation and stability of the enzyme-substrate complex, which is merely the first step in a cascade of events. The stability of the acyl-enzyme intermediate as well as the velocity of deacylation are important factors in defining the S/H ratio, but cannot be predicted from our modelling studies. For example, during deacylation, unforeseen steric hindrance might occur between the acyl donor residue and the groove of the active centre in papain. Furthermore, the type of nucleophile that attacks the enzyme-acyl intermediate is not taken into account in our modelling experiments.

Table 2.1 Various Z-X_{AA}- OGp esters tested experimentally[a]



				background	*enzymatic*	
Entry	**Amino acid**	**Time (min)**	**Conv. (%)**	**Z-X_{AA}-OH (%)**	**Z-X_{AA}-Phe-NH_2 (%)**	**Z-X_{AA}-OH (%)**
1	L-Ala	2	100	1.9	77.5	20.6
2	L-Arg	1	100	3.4	45.9[b]	17.2
3	L-Asn	25	100	1.4	6.1[c]	30.2
4	Gly	20	100	5.6	92.0	2.4
5	L-Glu	15	100	3.8	68.5	27.7
6	L-Ile	90	100	-	24.5[d]	12.2
7	L-Phe	25	100	3.4	29.8	66.8
8	L-Thr	20	100	4.3	90.7	5.0
9	L-Pro	180	5	5.0	-	-
10	D-Ala	180	82	13.5	11.6	5.4
11	β-Ala	180	39	7.4	31.6	-

[a] Conditions: 2 mM Z-X_{AA}-OGp, 15 mM H-Phe-NH_2, 0.2 M HEPES buffer pH 8.0, 0.2 M NaCl, 20 mM $CaCl_2$, 10% (v/v) DMF

[b] Non-enzymatic piperidinone side product formation (33.5%)

[c] Non-enzymatic succinimide side product formation (62.3%)

[d] Dipeptide product precipitated during reaction, estimated yield is 87.8%

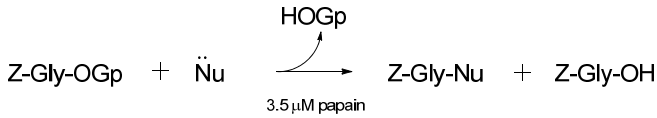
Despite the aforementioned shortcomings, it is clear that the high-ranking amino acids require the shortest reaction times, and that the reaction time generally increases with a decreasing predicted stability. However, phenylalanine does not seem to fit the correlation, with a much faster experimentally determined reaction time than suggested by its ranking. Apparently the hydrophobic pocket that determines the outcome of the MD simulation is of less influence on substrate mimetic binding than expected. Threonine also performed much better than predicted, based on our simulations. To explain this observation, we evaluated the different rotamers accessible to the threonine side chain. In the docking experiment, it was positioned in the *trans* χ_1 rotamer, which is not the typically preferred rotamer for threonine.[21] Therefore, we also assessed the stability of the more preferred *gauche* χ_1 rotamers and found that the *g* rotamer indeed resulted in a more stable complex, with an oxyanion hole-substrate hydrogen bond present for 41% of the simulation time. This is comparable to what is observed in Figure 2.2 for the other β-branched amino acids isoleucine and valine, which were both also predicted to bind in their preferred *g*- and *trans* rotamers.

It would not have been possible to decide *a priori* what percentage of hydrogen bond formation would be sufficient for activity. With the experimental results in hand, we can see that only a minor percentage of hydrogen bond formation and proximity is already enough to lead to enzymatic activity. Basically every experimentally tested amino acid in Table 2.1, except for proline, is accepted by papain. Even the two unnatural amino acids D-alanine and β-alanine, correctly predicted to react considerably slower, do react. The difference between predicted and experimentally determined reactivity may be taken as an indication that papain is in reality more flexible than modelling suggests.

2.2.5 Probing the influence of the amino acid in the nucleophile – acyl acceptor

Contrary to the acyl donor amino acids, the scope of the nucleophile was not computationally studied. As mentioned previously, papain displays broad specificity as a result of its wide-open peptide binding groove, which allows for many interactions upon substrate binding. The acyl donor can be modelled in this network by utilising known interactions with the oxyanion hole, but no such requirement is known for the incoming nucleophile. Hence, the influence of the acyl acceptor on the enzyme-specific activation was only probed experimentally with a restricted set of amino acid derivatives (Table 2).

Table 2.2 Various nucleophiles esters tested experimentally[a]



				background	*enzymatic*	
Entry	**Nucleophile**	**Time (min)**	**Conv. (%)**	**Z-Gly-OH (%)**	**Z-Gly-Nu-NH$_2$ (%)**	**Z-Gly-OH (%)**
1	D-Phe-NH$_2$	20	100	2.8	22.9	74.3
2	L-Phe-O^{t}Bu	25	100	3.8	11.8[b]	-
3	L-Tyr-NH$_2$	25	100	6.2	91.3	2.5
4	L-Tyr-*p*NA	25	100	5.1	90.6	4.3
5	L-Ala-NH$_2$	20	100	5.3	87.0	7.7
6	L-Ala-*p*NA	20	100	4.0	96.0	-
7	L-Ser-*p*NA	30	100	6.0	94.0	-
8	L-Pro-*p*NA	30	100	9.7	-	90.3
9	-	20	100	5.2	-	94.8

[a] Conditions: 2 mM Z-Gly-OGp, 15 mM Nu, 0.2 M HEPES buffer pH 8.0, 0.2 M NaCl, 20 mM $CaCl_2$, 10% (v/v) DMF

[b] Dipeptide product precipitated

Switching the configuration from L-Phe-NH_2 to D-Phe-NH_2 shifted the enzymatic reaction from primarily synthesis to mainly hydrolysis (entry 1). The exchange of an amide for a *tert*-butyl ester did not really affect the reaction time, although the dipeptide product precipitated (entry 2). The *p*NA-amide of phenylalanine did not dissolve under these reaction conditions and was therefore excluded from the set. As alternatives, the Tyr-NH_2 and Tyr-*p*NA pair was tested for the influence of the *p*NA group, prior to testing the less visible amino acids with this chromophore attached. The results were very much comparable, with or without *p*NA, also for the small amino acids alanine and serine. Proline, a worse nucleophile, only gave hydrolysis and that at a comparable rate to the one seen without an additional nucleophile being present.

2.3 Conclusion

In this chapter, we have shown that the OGp ester can be successfully applied in papain-catalysed dipeptide synthesis under aqueous conditions. Our docking studies, which were performed to increase our molecular understanding of the system, resulted in an unexpected binding mode for the Z-L-Ala-OGp ester, which is supported by experimentally observed structure-activity relationships. In contrast to the anticipated function of a substrate mimetic, the OGp ester showed an unprecedented enzyme-specific activation in papain. Prior to determining the scope of the acyl donor experimentally, we used a molecular-dynamics-simulations approach to prioritise 22 natural and unnatural amino acids for synthesis. The resulting ranking was in good agreement with the experimental data. A representative set of Z-X_{AA}-Phe-NH_2 dipeptides was obtained in moderate to excellent yields. The scope of the incoming nucleophile was relatively broad, ranging from the small Ala-NH_2 to the much larger Tyr-*p*NA. Since the OGp ester exhibits rather unexpected enzyme-specific activation, we hypothesise that it is not strictly required to have an arginine-like activating ester. Future investigations will be directed towards confirming the predicted binding mode of the OGp ester by using X-ray studies and to replacing it by a simpler moiety in order to render the described enzymatic approach to peptide-bond formation even more accessible.

2.4 Acknowledgements

S. C. B. Jans and H. I. V. Amatdjais-Groenen are kindly ackowledged for their contributions to the experimental part of this chapter. B. Zarzycka and dr. S. B. Nabuurs are gratefully ackowledged for performing the computational studies. S.B.N. is supported by the Netherlands Organization for Scientific Research (NWO) through a VENI grant (700.58.410).

2.5 Experimental Section

General procedure for the enzymatic reactions:

Enzymatic acyl transfer reactions were performed at 25 °C in a total volume of 375 µL containing 0.2 M HEPES buffer (pH 8.0), 0.2 M NaCl, 20 mM $CaCl_2$, 10% DMF and 2 mM *p*TSA as an internal standard. Stock solutions of Z-X_{AA}-OGp esters (50 mM) in DMF and nucleophiles (30 mM) in buffer were prepared. The final concentrations of acyl donor and acyl acceptor were 2 mM and 15 mM, respectively. The latter was calculated as free, N^{α}-unprotonated nucleophile concentration $[HN]_0$ according to the Henderson–Hasselbalch equation $[HN]_0 = [N]_0/(1 + 10^{pK-pH})$. Papain (4 mg) was activated before use by adding DTT (0.6 mg) and 1 mL 0.1 M phosphate buffer (pH 6.5) containing 2.5 mM EDTA and shaking for 10 minutes at 25 °C. This solution was stored on ice and was freshly prepared daily. Following thermal equilibration of assay mixtures, the enzymatic reactions were started by addition of papain at a final concentration of 3.5 µM. Blanc reactions were run in parallel, but milliQ was added instead of papain. From this control experiment the spontaneous ester hydrolysis could be determined, as well as non-enzymatic aminolysis of the acyl donor esters of which the latter could be ruled out. With regular intervals 20 µl aliquots were withdrawn and quenched with 20 µl glacial acetic acid. The reactions were monitored for 3 hours by HPLC and checked once more for changes in reaction mixture composition after 24 hours. The values reported are the average of at least two separate experiments. The identity of the formed peptide products was established by chemical synthesis of reference compounds and LC-MS.

HPLC-Analyses:

Samples were analysed by a Shimadzu LC 2010 analytical HPLC system equipped with a RP C18 column (Varian, Inertsil ODS-3, 5 µm, 150 × 4.6 mm) and eluted with various mixtures of acetonitrile/water containing 0.1% trifluoroacetic acid under isocratic and gradient conditions at flow rates of 1.0 mL·min^{-1}. The wavelength of detection was 254 nm. Product yields were calculated from peak areas of the substrate esters and the hydrolysis and aminolysis products.

Molecular modelling of the papain-peptide complex:

The molecular model of papain bound to the hexapeptide LLRLSL was constructed on the basis of the crystal structure of a papain-leupeptin complex (PDB entry 1POP) solved at 2.1 Å resolution.[16] This structure contains an LLR peptide bound to only the S-subsites. To gain more insight into peptide binding to the S'-subsites, a hybrid model was built using an LSL peptide fragment bound to the S'-subsites of another papain crystal structure (PDB entry 2CIO) solved at 1.5 Å resolution.[22] First, the two crystal structures were aligned using the MOTIF algorithm[23], after which the coordinates of the LSL peptide were transferred to the papain-leupeptin complex. Subsequently, a peptide bond between the LLR and LSL peptide fragment was manually added using the Yasara program[24] and finally the resulting complex was energy minimised using the Yasara2 forcefield.[25]

Molecular docking of papain substrates:

All molecular docking studies in papain were performed using the flexible docking program Fleksy.[15] The crystal structure of a papain-leupeptin complex (PDB entry 1POP) solved at 2.1 Å resolution[16] was used as the receptor structure. The structure was prepared for docking by removing leupeptin and all water molecules from the complex. Subsequently, hydrogen atoms were added to the structure and their positions were optimised using the Yasara program.[24] In the applied docking protocol only those docking poses were taken forward in which the scissile bond of the docked Z-X_{AA}-OGp compound aligned to the scissile bond of the natural peptide substrate. Otherwise, default parameters as described previously[15] were applied.

Molecular dynamics simulations:

The highest ranking docking poses obtained from the docking studies, were used as starting complex for molecular dynamics (MD) simulations using the Yasara program.[24] The complex was first solvated in a simulation cell two times 10 Å larger than the protein along each axis. The cell was neutralised by replacing water molecules with counter ions. The resulting system was first minimised with the Amber03 force field[26] using a 7.86 Å force cutoff and the Particle Mesh Ewald algorithm to treat long range electrostatic interactions. Simulated annealing was used (time step 2 fs, atom velocities scaled down by 0.9 every 10th step) until convergence was reached. Subsequently, 3 ns MD simulations were conducted at 298K for each of the 22 substrates with periodic boundary conditions and 1.25 fs time steps. Intermolecular forces were recalculated every two simulation steps and pressure control was employed to maintain a water density 0.997 g/cm^3.

2.6 References

[1] (a) K. Adermann, *Curr. Protein Pept. Sci.* **2005**, *6*, 205-206; (b) H. Korhonen, A. Pihlanto, *Curr. Pharm. Des.* **2003**, *9*, 1297-1308; (c) H. Korhonen, A. Pihlanto, *Curr. Pharm. Des.* **2007**, *13*, 829-843; (d) M. B. Roberfroid, *J. Nutr.* **2007**, *137*, 2493-2502; (e) B. Steffansen, C. U. Nielsen, S. Frokjaer, *Eur. J. Pharm. Biopharm.* **2005**, *60*, 241-245; (f) J. Torres-Llanez, B. Vallejo-Cordoba, A. F. González-Córdova, *Arch. Latinoam. Nutr.* **2005**, *55*, 111-117; (g) P. Vlieghe, V. Lisowski, J. Martinez, M. Khrestchatisky, *Drug Discovery Today* **2010**, *15*, 40-56.

[2] (a) R. Joshi, L. Meng, H. Eckstein, *Helv. Chim. Acta* **2008**, *91*, 983-992; (b) H. Xiang, G. Y. Xiang, Z. M. Lu, L. Guo, H. Eckstein, *Amino Acids* **2004**, *27*, 101–105.

[3] V. Schellenberger, H.-D. Jakubke, N. P. Zapevalova, Y. V. Mitin, *Biotechnol. Bioeng.* **1991**, *38*, 104-108.

[4] F. Bordusa, D. Ullmann, C. Elsner, H. D. Jakubke, *Angew. Chem. Int. Ed.* **1997**, *36*, 2473-2475.

[5] (a) K. Itoh, H. Sekizaki, E. Toyota, N. Fuijiwara, K. Tanizawa, *Bioorg. Chem.* **1996**, *24*, 59-68; (b) H. Sekizaki, K. Itoh, A. Shibuya, E. Toyota, K. Tanizawa, *Chem. Pharm. Bull.* **2007**, *55*, 1514—1517; (c) H. Sekizaki, K. Itoh, E. Toyota, K. Tanizawa, *Chem. Pharm. Bull.* **1998**, *46*, 846-849; (d) H. Sekizaki, K. Itoh, E. Toyota, K. Tanizawa, *Amino Acids* **1999**, *17*, 285-291.

[6] R. Günther, S. Thust, H. J. Hofmann, F. Bordusa, *Eur. J. Biochem.* **2000**, *267*, 3496-3501.

[7] R. Günther, A. Stein, F. Bordusa, *J. Org. Chem.* **2000**, *65*, 1672-1679.

[8] (a) R. Grünberg, I. Domgall, R. Günther, K. Rall, H. J. Hofmann, F. Bordusa, *Eur. J. Biochem.* **2000**, *267*, 7024-7030; (b) S. Xu, K. Rall, F. Bordusa, *J. Org. Chem.* **2001**, *66*, 1627-1632.

[9] (a) T. Fuchise, H. Kishimura, H. Sekizaki, Y. Nonami, G. Kanno, S. Klomklao, S. Benjakul, B.-S. Chun, *Food Chem.* **2009**, *116*, 611-616; (b) H. Sekizaki, E. Toyota, T. Fuchise, H. Zhou, Y. Noguchi, K. Horita, *Amino Acids* **2008**, *34*, 149-153; (c) E. Toyota, D. Iyaguchi, H. Sekizaki, M. Tateyama, K. K. S. Ng, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2009**, *65*, 717–723.

[10] (a) N. Wehofsky, M. Haensler, S. W. Kirbach, W. J.D., F. Bordusa, *Tetrahedron: Asymmetry* **2000**, *11*, 2421-2428; (b) N. Wehofsky, S. W. Kirbach, M. Haensler, J. D. Wissmanne, F. Bordusa, *Org. Lett.* **2000**, *2*, 2027-2030.

[11] N. Wehofsky, C. Wespe, V. Cerovsky, A. Pech, E. Hoess, R. Rudolph, F. Bordusa, *ChemBioChem* **2008**, *9*, 1493-1499.

[12] (a) N. Wehofsky, F. Bordusa, *FEBS Lett.* **1999**, *443*, 220-224; (b) N. Wehofsky, J. D. Wissmann, M. Alisch, F. Bordusa, *Biochim. Biophys. Acta* **2000**, *1479*, 114-122.

[13] M. Philipp, M. L. Bender, *Mol. Cell. Biochem.* **1983**, *51*, 5-32.

[14] S. Thust, B. Koksch, *J. Org. Chem.* **2003**, *68*, 2290-2296.

[15] S. B. Nabuurs, M. Wagener, J. d. Vlieg, *J. Med. Chem.* **2007**, *50*, 6507–6518.
[16] E. Schroder, C. Phillips, E. Garman, K. Harlos, C. Crawford, *FEBS Lett.* **1993**, *315*, 38-42.
[17] P. J. Berti, C. H. Faerman, A. C. Storer, *Biochemistry* **1991**, *30*, 1394-1402.
[18] M. Thormann, S. Thust, W. H. Hofmann, F. Bordusa, *Biochemistry* **1999**, *38*, 6056-6062.
[19] L. Zervas, T. T. Otani, M. Winitz, J. P. Greenstein, *J. Am. Chem. Soc.* **1959**, *81*, 2878-2884.
[20] T. Geiger, S. Clarke, *J. Biol. Chem.* **1987**, *262*, 785-794.
[21] R. J. Dunbrack, M. Karplus, *Nat Struct Biol.* **1994**, *1*, 334-340.
[22] M. S. Alphey, W. N. Hunter, *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2006**, *62*, 504-508.
[23] G. Vriend, C. Sander, *Proteins* **1991**, *11*, 52-58.
[24] E. Krieger, G. Koraimann, G. Vriend, *Proteins* **2002**, *47*, 393-402.
[25] (a) E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein, G. Vriend, *Proteins* **2004**, *57*, 678-683; (b) E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker, K. Karplus, *Proteins* **2009**, *77*, 114-122.
[26] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, *J. Comput. Chem.* **2003**, *24*, 1999-2012.

Chapter 3

Papain-specific activating esters in aqueous dipeptide synthesis

Enzymatic peptide synthesis has the potential to be a viable alternative for chemical peptide synthesis. Due to the increasing commercial interest in peptides, new and improved enzymatic synthesis methods are desirable. In recently developed enzymatic strategies such as substrate mimetic approaches and enzyme-specific activation, use of the guanidinophenyl ester (OGp) has been shown to suffer from some drawbacks. OGp-esters are sensitive to spontaneous chemical hydrolysis, expensive to synthesise and therefore not suitable for large-scale application. On the basis of earlier computational studies, we hypothesise that OGp might be replaceable by simpler ester groups to make the enzyme-specific activation approach to peptide bond formation more accessible. To this end a set of potentially activating esters (Z-Gly-*Act*) was designed, synthesised and evaluated. Both the benzyl (OBn) and the dimethylaminophenyl (ODmap) esters gave promising results. For these esters the scope of a model dipeptide synthesis reaction under aqueous conditions was investigated by varying the amino acid donor. The results were compared with those obtained from a previous study of Z-X_{AA}-OGp esters. Computational docking analysis of the set of esters was performed in order to provide insight into the differences in reactivities of all the potentially activating esters. Finally, selected ODmap- and OBn-activated esters were applied in the synthesis of two biologically active dipeptides on preparative scales.

This chapter has been published:
R. J. A. C. de Beer, B. Zarzycka, M. Mariman, H. I. V. Amatdjais-Groenen, M. J. Mulders, P. J. L. M. Quaedflieg, F. L. van Delft, S. B. Nabuurs, F. P. J. T. Rutjes, *ChemBioChem* **2012**, *13*, 1319-1326.

3.1 Introduction

The market share of peptide-based drugs is steadily increasing and will likely continue to do so, because a large number of therapeutic peptides are in clinical trials.[1] Other relevant applications of peptides include their use as nutrients[2] and cosmetic ingredients.[3] Large-scale peptide synthesis, however, is an expensive and time-consuming procedure, and this conflicts with the considerable commercial interest.[4]

Solid phase peptide synthesis (SPPS) is the most commonly used and versatile strategy for the synthesis of peptides in general.[5] The requirement of side-chain protection, the need to use excess of reagents, and the risk of racemisation are downsides of this method, but are to a large extent compensated by its capability to incorporate virtually any amino acid. A strategy commonly used in cases of larger peptides is to produce shorter fragments by SPPS and to couple them in solution to make the synthesis more convergent.[6] Alternatively, fragments with unprotected side chains can be coupled in aqueous solution by the chemoselective native chemical ligation method, which involves a *C*-terminal thioester and an *N*-terminal cysteine.[7] For large-scale production of long peptides (>50 amino acid residues) and protein sequences containing only natural amino acids, fermentation is currently most feasible, but this requires significant development efforts for each individual peptide or protein.[8] A relatively new and rather unexplored area is the use of enzymatic hydrolysis of readily available existing proteins to release specific peptides.[4a] An alternative use of these proteases, is their potential to create peptide bonds.[9] Advantages of such a chemoenzymatic approach include the high selectivity of the enzymes, minimal need for side-chain protection and absence of racemisation. On the other hand, the generally narrow substrate specificity of enzymes, in particular the exclusion of unnatural amino acids, severely restricts their universal application. In addition, undesired product hydrolysis when using proteases can be problematic, even though strategies have been developed to minimise this side reaction.[9] By employing the principle of substrate mimetics,[10] the challenge of inadequate substrate recognition can be overcome. In the substrate mimetics approach, the amino acid side chain – crucial for enzyme recognition – is transferred to the ester leaving group so that the enzymatic process becomes independent of the amino acid to be coupled. The guanidinophenyl (OGp) group, a mimetic of the arginine side chain, has been extensively applied in academic research with much success.[11] However, the OGp group is unsuitable for industrial application because it is expensive and difficult to work with due to the positively charged guanidino group.

We recently reported that the OGp ester can also be used in chemoenzymatic dipeptide synthesis under aqueous conditions with the protease papain.[12] In this case, the OGp moiety is likely not mimicking the natural substrate, because the guanidine group is predicted to bind to papain in a different manner than the arginine side chain. We named this phenomenon enzyme-specific activation, because no reaction takes place in the absence of the enzyme. From these findings, we hypothesised that it might not be essential to have an arginine-like activating ester for papain, as long as sufficient specific

interactions with the enzyme can be produced. A set of potentially enzyme-specific activating esters was therefore designed, synthesised and evaluated for effectiveness in papain-catalysed dipeptide formation.

3.2 Results and Discussion

3.2.1 Design of potentially enzyme-specific activating esters

In designing enzyme-specific activating esters for papain, several requirements had to be fulfilled. Firstly, sufficient functionalities (*e.g.*,hydrogen bond donors or acceptors) were incorporated to allow for interactions with the binding groove of papain. This is also important for gaining insight into the spatial, electronic and chemical requirements for papain specific activating esters. Secondly, the stability against spontaneous ester hydrolysis needs to be improved. Chemical background hydrolysis for OGp esters is substantial (~5% within 30 minutes)[12] and this is undesirable because it will increase when longer reaction times are required in more difficult couplings. Compared to phenolic esters, aliphatic or benzylic esters are considerably more stable. Thirdly, the new activating esters must preferably be cheap and easy to synthesise. A chemoenzymatic route – using Alcalase-CLEA,[13] for example – would be particularly attractive in this respect.

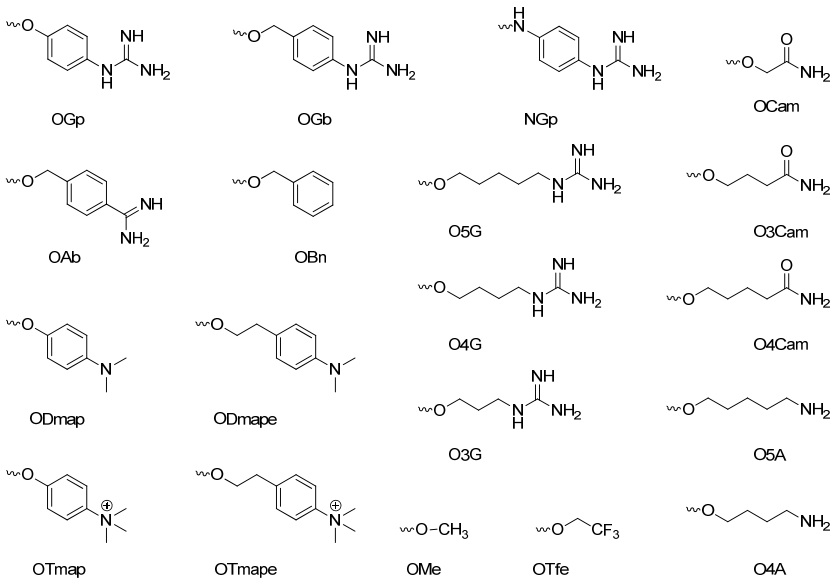
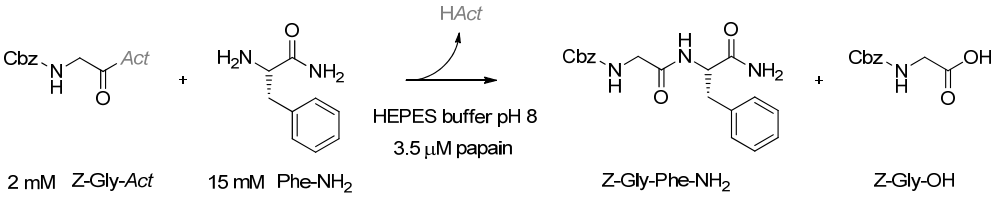


Figure 3.1 *Structures of potentially activating esters. OGp: p-guanidinophenyl ester; OGb: p-guanidinobenzyl ester; NGp: p-guanidinobenzyl amide; OAb: p-amidinobenzyl ester; OBn: benzyl ester; OCam: carbamoylmethyl ester; O3Cam: carbamoylpropyl ester; O4Cam: carbamoylbutyl ester; O3G: 3-guanidinopropyl ester; O4G: 4-guanidinobutyl ester; O5G: 5-guanidinopentyl ester; ODmap: p-(dimethylamino)phenyl ester; OTmap: p-(trimethylammonium)phenyl ester; ODmape: p-(dimethylamino)phenethyl ester; OTmape: p-(trimethylammonium)phenethyl ester; O4A: 4-aminobutyl ester; O5A: 5-aminopentyl ester; OMe: methyl ester; OTfe: 2,2,2-trifluoroethyl ester.*

N-protected amino acids can be converted into their alkyl or benzyl esters by this enzyme in near-anhydrous organic solvents. Furthermore, arylamides can be obtained by ester interconversion under the same conditions. The more labile phenyl esters are beyond the scope of this method.
These considerations led to a set of potentially activating moieties, as shown in Figure 3.1. These were chemically synthesised as the corresponding Z-Gly-OH esters. In addition, four activated esters that have been described in literature in combination with papain, *i.e.* OMe,[14] OCam,[15] OTfe[15] and OBn,[14a, 16] were also included for comparison. The synthesis details of all ester substrates can be found in Chapter 7.

3.2.2 Evaluation of potentially enzyme-specific activating esters

Next, the esters were evaluated for their ability to form the dipeptide Z-Gly-Phe-NH_2 with papain catalysis. The levels of conversion were determined by a previously described assay (Scheme 3.1).[12]

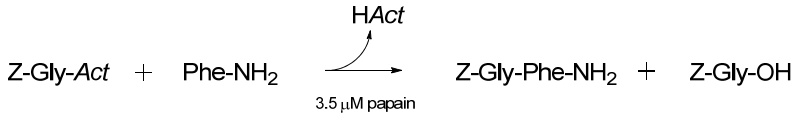


Scheme 3.1 *Enzymatic dipeptide synthesis with papain*

Table 3.1 shows either the time it takes to reach 100% conversion, or the levels of conversion after 3 hours. Background hydrolysis of the esters was determined from blank reactions in which no papain was present. The indicated percentages of enzymatic synthesis and hydrolysis remained constant over time as measured after 24 hours.
Ideally, the enzyme-specific activating group should give a fast reaction, a high percentage of synthesis and no hydrolysis. In none of the entries all these criteria are met at the same time, except for the phenolic ODmap ester (entry 6). However, the liberated dimethylaminophenol suffers from decomposition as deduced from the large amount of additional peaks in the HPLC chromatogram. The phenolic ester OTmap (entry 5) reacted comparably rapidly as OGp, but had a higher spontaneous hydrolysis rate. The introduction of two extra carbons (cf. ODmape and OTmape, entries 7 and 8, respectively) resulted in a significant increase in reaction time. The benzyl ester (OBn, entry 17) was superior to OGp in many respects, except for the time needed to reach full conversion of the starting material. The other benzylic esters, OGb and OAb (entries 2 and 4, respectively), reacted even more slowly and showed different percentages of enzymatic hydrolysis, which is remarkable in view of their similarity. The introduction of an amide bond (entry 3) resulted in complete inactivity. Of the set of aliphatic potentially activating esters, only the activated OCam and OTfe esters (entries 14 and 18, respectively) reacted

reasonably rapidly. As anticipated from their chemical properties, background hydrolysis was low for all aliphatic esters.

Table 3.1 Various Z-Gly-*Act* esters tested experimentally[a]



				background	*enzymatic*	
Entry	***Act***	**Time (min)**	**Conv. (%)**	**Z-Gly-OH (%)**	**Z-Gly-Phe-NH_2 (%)**	**Z-Gly-OH (%)**
1	OGp	20	100	5.6	92.0	2.4
2	OGb	180	97	3.3	92.5	1.2
3	NGp	180	-	-	-	-
4	OAb	180	99	2.2	86.5	10.3
5	OTmap	15	100	8.4	91.6	0
6	ODmap	10	100	1.2	98.6	0.2
7	OTmape	180	61	0.4	58.1	2.5
8	ODmape	180	21	0.1	18.9	2.0
9	O3G	180	86	1.8	68.6	15.6
10	O4G	180	83	1.6	66.8	14.6
11	O5G	180	66	1.2	54.7	10.1
12	O4A	180	69	2.5	53.9	12.6
13	O5A	180	70	0.6	56.7	12.7
14	OCam	15	100	1.8	89.9	8.3
15	O3Cam	180	80	1.1	67.8	11.1
16	O4Cam	180	97	0.2	81.9	14.9
17	OBn	45	100	0	97.5	2.5
18	OTfe	15	100	1.8	91.5	6.7
19	OMe	180	70	1.5	58.8	9.7

[a] Conditions: 2 mM Z-Gly-*Act*, 15 mM H-Phe-NH_2, 3.5 μM papain, 0.2 M HEPES buffer pH 8.0, 0.2 M NaCl, 20 mM $CaCl_2$, 10% (v/v) DMF

The enzymatic hydrolysis levels, on the other hand, were much higher in all cases. These results are in line with earlier studies showing that the Z-Gly esters of OMe,[14b] OBn,[16] OCam[14a] and OTfe[15] can be used in combination with papain. Because of the variety in documented reaction conditions – in organic solvent, in frozen aqueous solutions, varying equivalents, immobilised papain – it was, however, difficult to compare them properly. The fact that these reactions could not just be performed in buffer is consistent with the relatively high enzymatic hydrolysis values we find for these esters.

Although the enzymatic reactions proceed through the the same acyl-enzyme intermediate (Z-Gly-papain), differences in S/H ratio are observed. This might be because the first step (acylation of the enzyme) is not the rate-determining step for each ester. Another explanation might be that the leaving groups have varying affinities for the enzyme and therefore influence the attack of the nucleophile (water or H-Phe-NH_2) to larger or smaller extents.

3.2.3 Docking of activating esters into the active site of papain

We recently developed a molecular model of the interactions of Z-X_{AA}-OGp esters with the active site of papain.[12] A similar molecular modelling study was performed for each of the remaining 18 substrates using the flexible docking programme Fleksy.[17] Results were visualised and analysed by use of the YASARA programme. Several representative examples are shown in Figure 3.2 and discussed below.

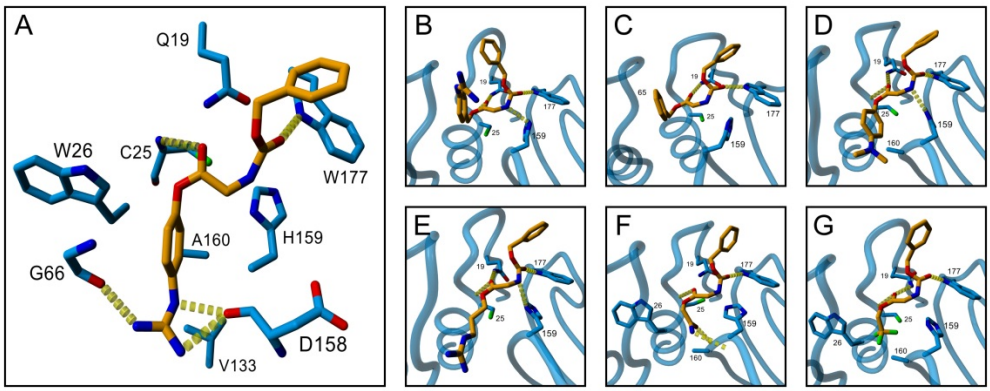


Figure 3.2 *Molecular modelling of a selection of esters in papain. Hydrogen bonding interactions to functionally important amino acids in the papain active site are shown for (A) Z-Gly-OGp, (B) Z-Gly-OGb, (C) Z-Gly-OBn, (D) Z-Gly-ODmap, (E) Z-Gly-O3G, (F) Z-Gly-OCam and (G) Z-Gly-OTfe.*

Z-Gly-OGp (Figure 3.2 A) fits tightly in the papain active site, which is located in a groove between two domains.[12, 18] The nucleophilic cysteine is as such positioned to attack the carbonyl carbon, whereas the carbonyl oxygen is situated in the oxyanion hole, which is made up of the side chain NH_2 group of Gln19 and the backbone amide of Cys25. The guanidinophenyl group is positioned in the groove, participating in several hydrogen bonding interactions.

In general, all the compounds discussed have the same *N*- to-*C* directionality as Z-Gly-OGp and fit the oxyanion hole. The most noticeable difference between Z-Gly-OGp and Z-Gly-OGb (Figure 3.2 B) is that the guanidinobenzyl group is now pointing out of the groove as a consequence of the extra carbon atom. As a result, the hydrogen bonding interactions with the groove are lost, which might explain the significant change in activity. Similarly, in Z-Gly-OBn (Figure 3.2 C) the benzyl group is also oriented towards the solvent, but it can favourably interact with Gly65 at the edge of the groove. Z-Gly-ODmap (Figure 3.2 D), a

phenolic ester, is nicely located in the groove and interacts with Tyr67 and Val133. Although the aliphatic Z-Gly-O3G (Figure 3.2 E) seems to be reasonably well situated in the groove, its reaction time is significantly longer than those of the phenolic esters. The small OCam moiety (Figure 3.2 F) participates in additional interactions with Trp26, Asp158 and Ala160 and fills the lower part of the peptide groove. High reactivity for OCam esters with α-chymotrypsin had already been reported to be attributable to orientation effects of hydrogen bonding interactions between the enzyme and the substrate.[19] Similarly, the OTfe group (Figure 3.2 G) exactly fills a hydrophobic cavity in the peptide groove, interacting with Trp26, Gly65 and Ala160.

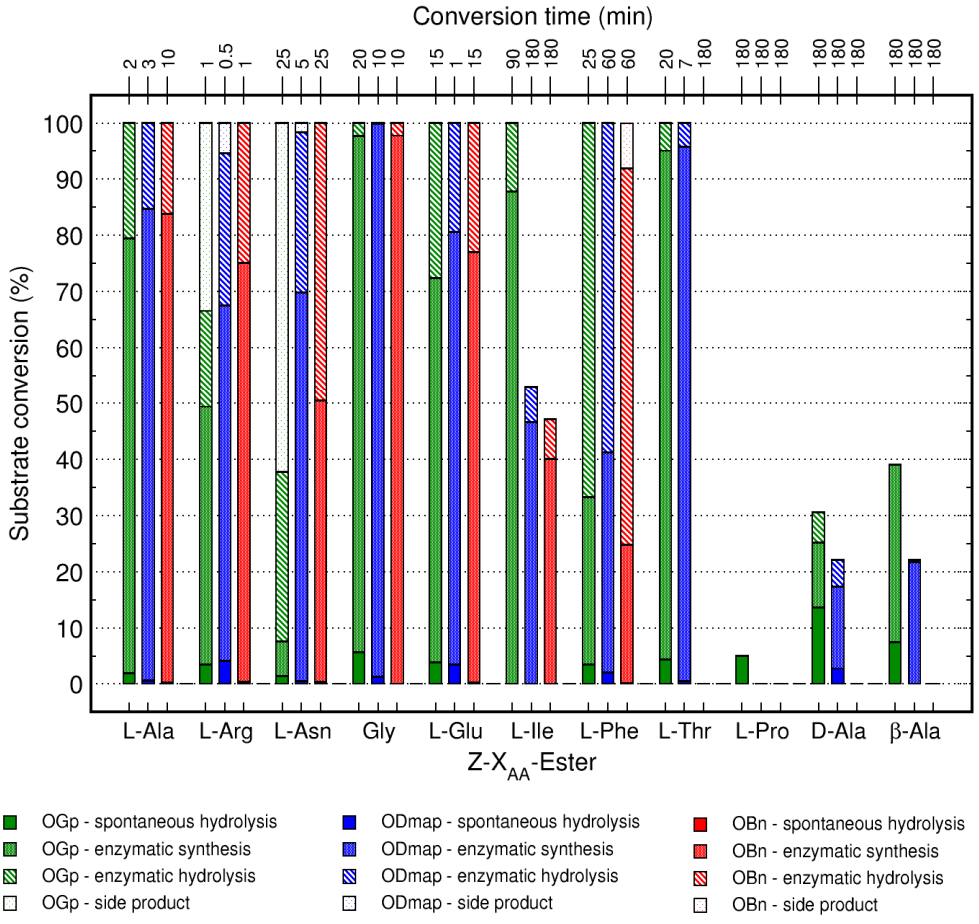


Figure 3.3 *Product distributions for dipeptide synthesis catalysed by papain for three esters (OGp, ODmap and OBn). Results are clustered per amino acid. Conditions: 2 mM Z-X_{AA}-Act, 15 mM H-Phe-NH_2, 3.5 μM papain, 0.2 M HEPES buffer (pH 8.0), 0.2 M NaCl, 10% (v/v) DMF. For reactions with Z-X_{AA}-OBn a double quantities of papain were used (7.0 μM). The dipeptide products Z-IF-NH_2 and Z-NF-NH_2 precipitated during the reaction, so the yield was estimated, Z-RF-NH_2 underwent secondary hydrolysis after 1 min and Z-FF-NH_2 was converted into Z-FF-OH in the OBn case.*[20]

These results imply that besides a good fit in the active site with sufficient interactions to the peptide groove and proper stabilisation by the oxyanion hole, the leaving group ability of the activating ester is also of significant importance. Clearly, the relatively poor leaving group ability of the aliphatic esters cannot be compensated for by the enzyme, whereas the differences between the much better phenolic leaving groups can be explained on the basis of their interactions with papain. The results also show that the good leaving groups OCam and OTfe are to some extent even further activated by the enzyme. The outcome with the benzyl ester is remarkable in our view, because the inferior leaving group ability is largely overcome by enzyme-specific activation.

3.2.4 Scope of the amino acid donor in Z-X_{AA}-OBn and Z-X_{AA}-ODmap

The high levels of conversion, combined with low hydrolysis rates, observed with the readily accessible ODmap and OBn esters led us to select these two for further exploration. In particular, in order to probe the more general applicability of the concept of enzyme-specific activation, we investigated whether the relatively broad range of amino acid donors as determined for OGp[12] would be maintained with ODmap and OBn, in order to probe. Analogous to the procedure followed for OGp esters, Z-X_{AA}-ODmap and Z-X_{AA}-OBn esters were synthesised from a representative set of amino acids with appropriate acid-labile side-chain protection and evaluated in the enzymatic assay. In the case of OBn esters, the amount of papain was doubled to compensate for the slower reaction rate that had been observed with Z-Gly-OBn. The identities of the products in the enzymatic reactions were confirmed by chemical synthesis of reference compounds and LC-MS analysis. The distribution of background hydrolysis, enzymatic synthesis, enzymatic hydrolysis, and in some cases side product formation for each of the studied esters is shown in Figure 3.3. The background hydrolysis was determined from a blank reaction where no papain was present. Above each bar, either the time is indicated that is required to reach 100% conversion, or 3 hours, after which regular monitoring was stopped. The indicated product distribution remained constant over time, as measured after 24 hours, unless stated otherwise.

For ease of comparison, the results for Z-X_{AA}-OGp[12] are also included in the diagram, and the data are clustered per amino acid. Some trends can be distinguished; generally, the percentage of enzymatic synthesis (■) over hydrolysis (▨) is comparable for each individual amino acid independent of the ester. This was to be expected because the binding pocket for the amino acid did not change. A remarkable exception in this respect is the complete inactivity of Z-L-Thr-OBn towards papain, which may be due to steric effects.

The spontaneous background hydrolysis (■) decreased on going from the OGp ester to the ODmap and OBn esters, which correlates well with the decreasing leaving group ability of these esters. The non-enzymatic cyclic side product formation (⊡; a piperidone from Arg and a succinimide from Asn) dropped drastically in this order for the same reason. All of the OBn esters reacted considerably more slowly than the OGp and ODmap esters, despite the double quantities of papain. This difference in reactivity became even more

apparent for the more challenging amino acids Z-L-Ile-OH, Z-D-Ala-OH and Z-β-Ala-OH. Increasing the reaction time for the corresponding OBn esters did not result in significantly higher conversions, but dipeptide synthesis occurred when the quantities of papain were simultaneously increased by a further factor of ten, (Table 3.2).

Table 3.2 Various Z-$_{XAA}$-OBn esters tested experimentally with a high papain concentration[a]

				background	*enzymatic*	
Entry	**Amino acid**	**Time (h)**	**Conv. (%)**	**Z-X_{AA}-OH (%)**	**Z-X_{AA}-Phe-NH_2 (%)**	**Z-X_{AA}-OH (%)**
1	L-Ile	6	100	-	30.9 (85.5)[b]	14.5
2	D-Ala	10	100	0.8	51.7	47.5
3	β-Ala	24	26	-	25.4	0.6

[a] Conditions: 2 mM Z-X_{AA}-OBn, 15 mM H-Phe-NH_2, 70 µM papain, 0.2 M HEPES buffer pH 8.0, 0.2 M NaCl, 20 mM $CaCl_2$, 10% (v/v) DMF

[b] Dipeptide product precipitates during reaction, estimated yield is given in parentheses

The scope of the amino acid donor is comparably broad for ODmap and OGp esters. The OBn ester also enables the papain catalysed synthesis of various dipeptides, provided that the amount of enzyme is increased. This seems an acceptable trade-off given the fact that OBn esters are readily accessible, even with an enzymatic approach.

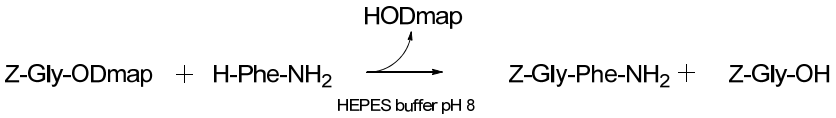
3.2.5 Synthesis of a selected dipeptide on preparative scale

The applicability of the enzyme-specific activation strategy was illustrated by means of preparative scale syntheses of two biologically active peptides: Z-Gly-Phe-NH_2 with Z-Gly-ODmap as activating ester, and H-Glu-Trp-OH starting from enzymatically synthesised Z-L-Glu-OBn. The first dipeptide is a substrate for metalloendoproteases blocking the exocytotic release of histamine and catecholamines from mast cells and adrenal chromaffin cells.[21] Furthermore, it interferes with insulin processing and inhibits glucose transport in adipocytes.[22] The second is also known as thymogen, oglufanide, or IM862, a naturally occuring immunomodulator, which is a potent anti-angiogenic agent and normallises the immune system function of immunocompromised individuals.[1a, 23]

Firstly, the small-scale experiment (Table 3.3, entry 1) was repeated with 100 mg acyl donor instead of 0.25 mg, resulting in a longer reaction time and a slightly increased amount of hydrolysis as determined by HPLC analysis of a sample from the reaction mixture (entry 2). The desired dipeptide was isolated in 74% yield after column chromatography. Next, the excess of H-Phe-NH_2 was decreased to arrive at equimolar amounts of acyl donor and acceptor (entry 3). At this point, additional DMF was required to keep all compounds in solution, and the papain concentration was increased tenfold to compensate for the anticipated loss of activity. These conditions resulted in a faster reaction, a somewhat better S/H ratio and an isolated yield of 79%. Finally, this reaction was performed again on gram scale (entry 4). Despite the ten times higher papain

concentration, the reaction was not completed after 75 minutes, but the product was purified nevertheless. The isolated yield of 74% was equal to the yield calculated from the level of conversion indicated by HPLC.

Table 3.3 Reaction conditions and yields preparative scale synthesis of Z-Gly-Phe-NH_2



Entry	Scale (mg)	D/A[a] (mM)	DMF % (v/v)	Papain (μM)	Time (min)	Conv.[b] (%)	S/H ratio[c] (%)	Yield[d] (%)
1	0.25	2:15	10	3.5	10	100	98:2	n.a.
2	100	2:15	10	3.5	120	97	85:12	74
3	100	15:15	20	35	60	99	90:9	79
4	1000	15:15	20	350	75	89	74:15	74

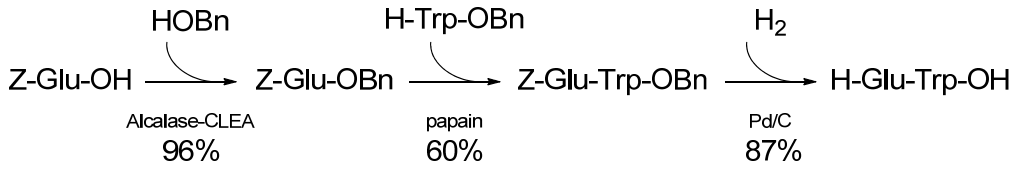
[a] Donor to acceptor ratio Z-Gly-ODmap : H-Phe-NH_2

[b] According to HPLC

[c] Synthesis over hydrolysis ratio Z-Gly-Phe-NH_2 : Z-Gly-OH

[d] Isolated yield of dipeptide after purification by flash chromatography

The synthesis of the second dipeptide was envisioned as depicted in Scheme 3.2. The chemoenzymatic method developed by Nuijens was employed to benzylate Z-Glu-OH selectively at the α-position and the activating ester was obtained in 96% yield.[13a] This synthesis clearly emphasises the advantage of OBn as activating ester over phenolic esters, which cannot be enzymatically prepared.. H-Trp-OBn was selected as the nucleophile to enable the simultaneous removal of both protecting groups, but with the risk of repeated coupling due to the continued presence of the OBn group, which is recognised by papain.
Test experiments on preparative scale indicated that multiple addition of H-Trp-OBn did indeed occur, but that it could be minimised by stopping the reaction in time. These experiments also showed that the reaction proceeds much more rapidly but with a worse S/H ratio than in the case of the previous dipeptide. This led us to conduct a 500 mg scale experiment with a 1:1 donor/acceptor ratio and a papain concentration of 8.8 μM. Under these conditions, HPLC analysis showed full conversion within 45 min and an S/H ratio of 66:27; additionally, a 7% yield of tripeptide was formed. The desired dipeptide was isolated in pure form in 60% yield after column chromatography. Subjection of the protected dipeptide to hydrogenolysis conditions provided biologically active H-Glu-Trp-OH in 87% yield.



Scheme 3.2. *Synthesis of bioactive dipeptide H-Glu-Trp-OH. Reaction conditions step 1: Z-Glu-OH, BnOH, Alcalase-CLEA, THF, mol sieves (3 Å), 150 rpm, 37 °C; step 2: 5 mM Z-Glu-OBn, 5 mM H-Trp-OBn, 8.8 μM papain, 0.2 M Hepes buffer pH 8.0, 0.2 M NaCl, 20 mM $CaCl_2$, 15% (v/v) DMF, 25 °C; step 3: H-Cube, 10% Pd/C, 10 bar H_2, 1 mL/min, rt.*

3.3 Conclusion

In this chapter, we have shown that papain-catalysed dipeptide synthesis can also be performed successfully with activating esters that do not resemble arginine. Out of a set of potentially activating esters, both the ODmap and OBn esters appeared to be suitable replacements for OGp, with the added benefits of (1) being simpler in structure and therefore cheaper and easy to synthesise, and (2) giving smaller amounts of undesired hydrolysis product. The scopes of the ODmap and OBn acyl donors were compared with that of the OGp esters and it was found that the S/H ratios are variable for the different amino acids, but rather independent of the ester involved. The reactions with OBn esters required increased quantities of papain to afford similar reaction rates. The applicability of the ODmap ester was validated in preparative-scale syntheses of two biologically active dipeptides.

The computational model of papain, which was originally built for OGp, was effectively applied in a docking study to get insight into the differences in reactivity of all the potentially activating esters. This exercise showed that besides a proper fit in the active site, the leaving group ability is also important. To shed light on this intricate relationship, further studies, both computational and experimental, are currently underway. The balance between a good leaving group ability to enable the enzymatic reaction, a stable ester that does not undergo rapid spontaneous hydrolysis and the extent to which the ester is activated by the enzyme seems to be delicate.

3.4 Acknowledgements

M. Mariman and M. J. Mulders are kindly ackowledged for their contributions to the synthesis of the activating esters. Timo Nuijens (DSM) and Martin Broxterman are acknowledged for their help with the enzymatic synthesis of Z-Glu-OBn. H. I. V. Amatdjais-Groenen is gratefully ackowledged for the preperative scale experiments. B. Zarzycka and dr. S. B. Nabuurs are gratefully ackowledged for performing the computational studies.

3.5 Experimental section

Synthesis: see Chapter 7 for a detailed description of the synthetic procedures and product characterisation.

Computational studies: a detailed description of the computational analyses can be found in a previous article on this subject.[12]

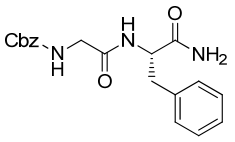
Enzymatic acyl transfer reaction: see experimental § 2.5.

Enzymatic synthesis on preparative scale:

To a round-bottomed flask was added the acyl donor, the acyl acceptor and pTSA as an internal standard. This was disolved in a small amount of DMF to which was added HEPES buffer with a final concentration of 0.2 M, pH 8 and containing 0.2 M NaCl and 20 mM CaCl2. Finally, papain, activated with DTT in phosphate buffer (0.1 M, pH 6.5) containing EDTA (2.5 mM), was added to start the reaction. The mixture was stirred at 25 °C until total conversion of the starting material was observed, which was determined by HPLC. Then, the reaction mixture was extracted with EtOAc (3×). The combined organic layers were washed with brine, dried over sodium sulfate and evaporated to dryness.

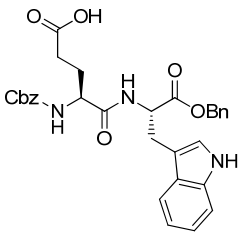
Acyl donor	(mg)	Acyl acceptor	(mg)	*p*TSA (mg)	DMF (mL)	buffer (mL)	milliQ (mL)	papain (mg)
Z-Gly-ODmap	99	H-Phe-NH_2	443	57	15	75	60	12
Z-Gly-ODmap	99	H-Phe-NH_2	59	7.6	4	10	6	16
Z-Gly-ODmap	985	H-Phe-NH_2	59	76	40	100	60	160
Z-Glu-OBn	557	H-Trp-OBn	496	114	45	150	105	60

***N*$^{\alpha}$-Cbz-Glycine-phenylalanine amide (Z-Gly-Phe-NH_2)**



The product was obtained as a white solid after purification by column chromatography (MeOH in CH_2Cl_2, 2→5%). R_f 0.45 (10% MeOH in CH_2Cl_2). Mp 143.0 °C. $[\alpha]_D^{20}$ +7.1 (*c* 2.60, MeOH). IR (film) 3460, 3299, 3062, 3032, 1712, 1681, 1644, 1563, 1529, 1284, 1250, 1219, 1172, 1045, 992, 728, 697 cm^{-1}. ^1H NMR (CD3CN, 400 MHz): δ 7.42-7.18 (m, 10H), 6.88 (d, *J* = 7.7 Hz, NH), 6.44 (br s, NH), 5.98 (br t, NH), 5.83 (br s, NH), 5.07 (s, 2H), 4.52 (ddd, *J* = 5.2, 8.3, 8.3 Hz, 1H), 3.67 (dd, *J* = 6.1. 17.0 Hz, 1H), 3.62 (dd, *J* = 6.0. 16.9 Hz, 1H), 3.13 (dd, *J* = 5.1, 14.0 Hz, 1H), 2.89 (dd, *J* = 8.5, 14.0 Hz, 1H). ^{13}C NMR (CD_3CN, 75 MHz): δ 173.8, 170.1, 157.7, 138.4, 137.9, 130.2, 129.4, 129.2, 128.9, 128.7, 127.5, 67.3, 54.9, 44.9, 38.2. HRMS (ESI) *m/z* calcd for $C_{19}H_{21}N_3NaO_4$ $(M+Na)^+$: 378.1430, found: 378.1443.

***N*$^{\alpha}$-Cbz-Glutamate-Tryptophan benzyl ester (Z-Glu-Trp-OBn)**



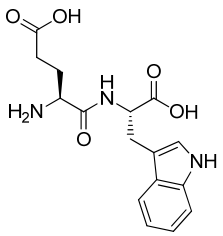
The product was obtained as a white solid after purification by column chromatography (AcOH/MeOH/ EtOAc 1:5:94) and lyophilisation from dioxane. R_f 0.50 (AcOH/MeOH/EtOAc 1:25:474). Mp 48.7 °C. $[\alpha]_D^{20}$ −13.8 (*c* 1.00, MeOH). IR (film) 3312, 3058, 3031, 2963, 2918, 1714, 1664, 1528, 1455, 1342, 1213, 743, 698 cm^{-1}. ^1H NMR (CD_3CN, 400 MHz): δ 9.17 (s, NH), 7.51 (d, *J* = 7.9 Hz, NH), 7.41-7.19 (m, 11H), 7.15-7.10 (m, 1H), 7.08-7.01 (m, 3H), 5.99 (d, *J* = 7.4 Hz, NH), 5.09-4.99 (m, 4H), 4.75-4.69 (m, 1H), 4.13-4.05 (m, 1H), 3.29-3.15 (m, 2H), 2.30 (t, *J* = 7.5 Hz, 2H), 1.82-1.71 (m, 2H). ^{13}C NMR (CD_3CN, 75 MHz): δ 174.7, 172.4, 172.3, 157.1, 137.3, 136.8, 129.4, 129.1, 129.0, 128.8, 128.6, 128.3, 124.7, 122.5, 120.0,

119.2, 118.2, 112.2, 110.3, 67.6, 67.2, 55.1, 54.2, 30.4, 28.2, 28.1. HRMS (ESI) *m/z* calcd for $C_{31}H_{31}N_3O_7$ $(M+H)^+$: 558.2240, found: 558.2240.

Chemoenzymatic synthesis of Z-Glu-OBn:

Before use, Alcalase-CLEA (3 g, Type OM, CLEA-Technologies, 580 U/g) was suspended in ^{t}BuOH (100 mL) and crushed with a spatula. After filtration, the enzyme was resuspended in MTBE (50 mL) followed by filtration. Large enzyme particles were removed by a sieve (0.5 mm pore size). 3 Å Molecular sieves (Acros, 8 to 12 mesh) were activated (200°C under vacuum overnight), crushed and sieved (0.5 mm pore size) to remove large particles.
Alcalase-CLEA (509 mg) was added to a solution of Z- L-Glu-OH (977 mg, 3.48 mmol, 1 equiv) in THF (9 mL) and benzylalcohol (1 mL, 9.66 mmol, 2.8 equiv). The mixture was shaken at 37 °C at 150 rpm for 16 h in the presence of 3 Å molecular sieves. After filtration, the enzyme was washed by resuspension in THF (3 × 20 mL) followed by filtration. The combined organic layers were concentrated *in vacuo* and the resulting oil was purified by column chromatography using aminomethyl resin (TFA in CH_2Cl_2, 0→2.5%). The solution was concentrated, TFA was removed by co-evaporation with toluene (2 × 20 mL) and $CHCl_3$ (2 × 20 mL) and the product was lyophilised from MeCN/H_2O 3:1 to give Z-Glu-OBn(1.24 g, 96%) as a white solid. Spectral data were in accordance with those reported in the supporting information for the chemically synthesized compound.

Glutamate-tryptophan (H-Glu-Trp-OH)



Z-Glu-Trp-OBn (121 mg, 0.22 mmol, 1 equiv) was dissolved in 20% H_2O in MeOH (12 mL) and hydrogenated with 10% Pd/C in the H-Cube (ThalesNano) at room temperature, 1 mL/min and 10 bar H_2 pressure. The product was obtained as a slightly pink solid after evaporation of the solvent *in vacuo* (63 mg, 87%). R_f 0.27 ($CHCl_3$/MeOH/NH_4OH 65:45:20). Mp 159.2 °C. $[\alpha]_D^{20}$ +22.4 (*c* 0.31, DMSO). IR (film) 3213, 3067, 1693, 1665, 1597, 1548, 1483, 1455, 1414, 1339, 740 cm^{-1}. ^1H NMR (CD_3OD/D_2O, 400 MHz): δ 7.64 (d, *J* = 7.8 Hz, 1H), 7.37 (d, *J* = 8.1 Hz, 1H), 7.18 (s, 1H), 7.11 (ddd, *J* = 1.1, 7.1, 8.1 Hz, 1H), 7.04 (ddd, *J* = 1.1, 7.1, 8.0 Hz, 1H), 4.61 (dd, *J* = 4.8, 8.7 Hz, 1H), 3.82 (dd, *J* = 5.2, 7.4 Hz, 1H), 3.39 (dd, *J* = 4.8, 14.7 Hz, 1H), 3.18 (dd, *J* = 8.8, 14.7 Hz, 1H), 2.52-2.36 (m, 2H), 2.16-1.95 (m, 2H). ^{13}C NMR (CD_3OD/D_2O, 75 MHz): δ 179.5, 177.5, 169.8, 137.7, 128.7, 124.7, 122.4, 119.8, 119.4, 112.4, 111.5, 56.9, 54.3, 33.3, 28.6, 28.4. HRMS (ESI) *m/z* calcd for $C_{16}H_{20}N_3O_5$ $(M+H)^+$: 334.1403, found: 334.1391.

3.6 References

[1] (a) P. Vlieghe, V. Lisowski, J. Martinez, M. Khrestchatisky, *Drug Discovery Today* **2010**, *15*, 40-56; (b) D. Hans, P. R. Young, D. P. Fairlie, *Med. Chem.* **2006**, *2*, 627-646; (c) R. E. W. Hancock, H.-G. Sahl, *Nat. Biotechnol.* **2006**, *24*, 1551-1557; (d) A. A. Zompra, A. S. Galanis, O. Werbitzky, F. Albericio, *Future Med. Chem.* **2009**, *1*, 361-377.

[2] (a) V. A. Tutel'yan, V. K. Khavinson, V. V. Malinin, *Bull. Exp. Biol. Med.* **2003**, *135*; (b) R. Hartmann, H. Meisel, *Curr. Opin. Biotechnol.* **2007**, *18*.

[3] (a) M. P. Lupo, A. L. Cole, *Dermatol. Ther.* **2007**, *20*, 343–349; (b) B. S. Oricha, *Afr. J. Pharm. Pharmacol.* **2010**, *4*, 127-129.

[4] (a) D. Agyei, M. K. Danquah, *Biotechnol. Adv.* **2011**, *29*; (b) V. Marx, *Chem. Eng. News* **2005**, *83*, 17-24; (c) L. Andersson, L. Blomberg, M. Flegel, L. Lepsa, L. Nilsson, M. Verlander, *Biopolymers (Pept. Sci.)* **2000**, *55*, 227–250.

[5] F. Albericio, *Curr. Opin. Chem. Biol.* **2004**, *8*, 211-221.

[6] J. Y. Lee, D. Bang, *Biopolymers (Pept. Sci.)* **2010**, *94*, 441-447.

[7] S. B. H. Kent, *Chem. Soc. Rev.* **2009**, *38*, 338-351.

[8] I. Gill, R. López-Fandiño, X. Jorba, E. N. Vulfson, *Enzyme Microb. Technol.* **1996**, *18*, 162-183.
[9] F. Guzmán, S. Barberis, A. Illanes, *J. Biotechnol.* **2007**, *10*, 279-314.
[10] F. Bordusa, *Chem. Rev.* **2002**, *102*, 4817-4867.
[11] F. Bordusa, *Curr. Protein Pept. Sci.* **2002**, *3*, 159-180.
[12] R. J. A. C. de Beer, B. Zarzycka, H. I. V. Amatdjais-Groenen, S. C. B. Jans, T. Nuijens, P. J. L. M. Quaedflieg, F. L. van Delft, S. B. Nabuurs, F. P. J. T. Rutjes, *ChemBioChem* **2011**, *12*, 2201-2207.
[13] (a) T. Nuijens, C. Cusan, J. A. W. Kruijtzer, D. T. S. Rijkers, R. M. J. Liskamp, P. J. L. M. Quaedflieg, *Synthesis* **2009**, *5*, 809–814; (b) T. Nuijens, C. Cusan, J. A. W. Kruijtzer, D. T. S. Rijkers, R. M. J. Liskamp, P. J. L. M. Quaedflieg, *J. Org. Chem.* **2009**, *74*, 5145–5150.
[14] (a) M. Fité, G. Alvaro, P. Clapés, J. López-Santin, M. D. Benaiges, G. Caminal, *Enzyme Microb. Technol.* **1998**, *23*, 199-203; (b) P. Stehle, H.-P. Bahsitta, B. Monter, P. Fürst, *Enzyme Microb. Technol.* **1990**, *12*, 56-60; (c) A. Lang, C. Hatscher, P. Kuhl, *Tetrahedron Lett.* **2007**, *48*, 3371-3374.
[15] S. M. A. Salam, K. Kagawa, T. Matsubara, K. Kawashiro, *Enzyme Microb. Technol.* **2008**, *43*, 537–543.
[16] S.-T. Chen, K.-T. Wang, *J. Chem. Soc., Chem. Commun.* **1988**, 327-328.
[17] S. B. Nabuurs, M. Wagener, J. de Vlieg, *J. Med. Chem.* **2007**, *50*, 6507–6518.
[18] E. Schroder, C. Phillips, E. Garman, K. Harlos, C. Crawford, *FEBS Lett.* **1993**, *315*, 38-42.
[19] M. Capellas, M. D. Benaiges, G. Caminal, G. Gonzalez, J. Lopez-Santin, P. Clapés, *Biotechnol. Bioeng.* **1996**, *50*, 700-708.
[20] This is in line with literature findings that the second amino acid (occupying the P_2 position) is important for the recognition by papain: (a) I. Schechter, A. Berger, *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157-162; Other amino acids that are recognised by papain in that position include Val and Tyr, but these were not present in our selection: (b) Y. Choe, F. Leonetti, D. C. Greenbaum, F. Lecaille, M. Bogyo, D. Brömme, J. A. Ellman, C. S. Craik, *J. Biol. Chem.* **2006**, *281*, 12824-12832
[21] D. I. Mundy, W. J. Strittmatter, *Cell* **1985**, *40*, 645-656.
[22] (a) L. P. Aiello, M. Wessling-Resnick, P. F. Pilch, *Biochemistry* **1986**, *25*, 3944-3950; (b) A. Jochen, P. Berhanu, *Biochem. Biophys. Res. Commun.* **1987**, *142*, 205-212.
[23] J. A. Bergeon, Y.-N. Chan, B. G. Charles, I. Toth, *Biopolymers* **2008**, *90*, 633-643.

Chapter 4

Enzyme-specific activation versus leaving group ability

Enzyme-specific activation and the substrate mimetics strategy are effective ways to circumvent the limited substrate recognition often encountered in protease-catalysed peptide synthesis. A key structural element in both approaches is the guanidinophenyl (OGp) ester, which enables important interactions for affinity and recognition by the enzyme. At least, this is usually the explanation given for its successful application. In this study we show that leaving group ability is of equal or even greater importance utilising experimental and computational methods, comprising (1) the synthesis of close analogues of OGp and their evaluation in a dipeptide synthesis assay with trypsin, (2) molecular docking studies that provided insight into the binding mode and (3) *ab initio* calculations to evaluate their electronic properties.

This chapter will be published:
R. J. A. C. de Beer, B. Bögels, G. Schaftenaar, B. Zarzycka, P. J. L. M. Quaedflieg, F. L. van Delft, S. B. Nabuurs, F. P. J. T. Rutjes, *Accepted in ChemBioChem* **2012**, doi 10.1002/cbic.201200227.

4.1 Introduction

These days enzymes are commonly used in organic synthesis.[1] The benefits of enzymatic reactions, including their generally excellent regio- and enantioselectivity, are widely recognised. Because of the usually mild reaction conditions, enzymatic conversions are often regarded as a green alternative to classic organic reactions. However, the limited substrate scope of many enzymes remains a big disadvantage.

One of the areas of application of enzymes is peptide synthesis. In this field, proteases are employed to form the peptide bonds (which they would natively hydrolyse) by exploiting the reversibility of each chemical reaction. A prerequisite for this enzymatic activity, irrespective of whether aqueous media or organic solvents are used, is that the amino acid in question is specifically recognised.[2] The latter problem of recognition can be circumvented by applying the substrate mimetics strategy as previously described for trypsin and other proteases.[3] The guanidinophenyl (OGp) ester, which in essence resembles the naturally recognised side chain of arginine, is claimed to serve as a recognition moiety for trypsin, thereby making recognition independent of the side chain of the amino acid and thus broadening the substrate scope. This approach is typically applied under aqueous conditions, but because OGp also functions as a leaving group, the commonly occurring issue of secondary hydrolysis is prevented, as the product formed becomes unrecognisable for the enzyme.

A similar solution to limited substrate acceptance was found for papain, *i.e.* enzyme-specific activation.[4] Based on docking studies, the OGp group is predicted to bind to the enzyme in a different orientation than its natural substrate arginine. Nevertheless, due to alternative recognition papain is able to catalyse dipeptide formation with unspecific amino acid residues. We noticed these versatile applications of the OGp moiety and wondered what the reason would be for these remarkable properties. Additional research into the replacement of OGp for simpler esters indicated that besides recognition and affinity for the enzyme, the leaving group ability may be an important factor.[5] In this study, the contribution of these components is investigated, using both experimental and computational methods. Several analogues of OGp were designed, synthesised and docked in trypsin to provide insight into the binding mode. Subsequently, their effectiveness in dipeptide formation was experimentally determined and an attempt was made to increase the activity by further variation of one of the analogues. An *ab initio* study provided insight into the electronic properties of the analogues under investigation.

4.2 Results & Discussion

4.2.1 Prediction of the binding mode of OGp analogues in trypsin

To distinguish between the effects of affinity for the enzyme and leaving group ability, we evaluated a set of close analogues of OGp that differed slightly in both properties. We opted for trypsin as the model system, because this enzyme is highly specific for arginine, in contrast to the aforementioned protease papain, which exhibits broad substrate specificity

with only a slight preference for arginine. Moreover, the catalytic mechanism of the serine protease trypsin,[6] also with respect to substrate mimetics,[7] is well known.
The analogues were designed in such a way that they closely resemble OGp in structure, while the leaving group character was varied considerably (Figure 4.1).

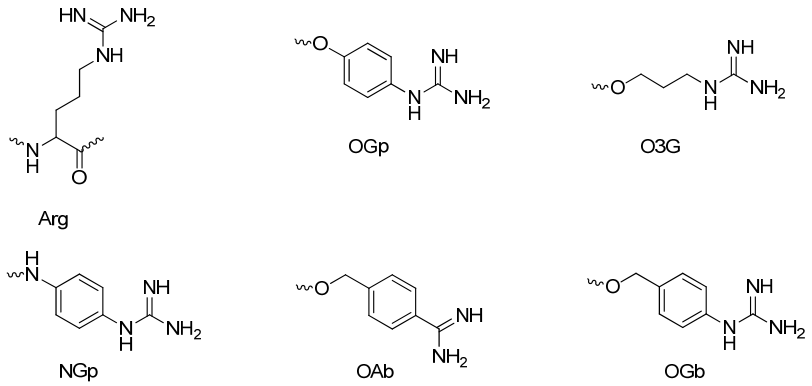


Figure 4.1 *Structures of OGp analogues*

Two benzylic variants (OAb and OGb) were synthesised as the corresponding Z-Gly-OH esters,[5] as well as an aliphatic analogue (O3G) and an OGp equivalent where the ester bond was replaced by an amide bond (NGp). The substrates were restricted to glycine esters in order to rule out any influence of the side chain in the coupling reactions. The anticipated analogy of these four compounds was evaluated using a computational docking study, with the flexible docking programme Fleksy.[8] The results were visualised and analysed by using the YASARA programme.[9]
The three dimensional structure of trypsin has been previously solved by crystallographic studies.[10] The secondary structure of this globular enzyme consists of β-structures organised into two densely packed hydrophobic barrels. Five subpockets (S_3 through S_2')[11] are important for specific binding of the substrate. A crucial interaction occurs with the carboxylate moiety of Asp189 at the bottom of subpocket S_1, which primarily determines the specificity of trypsin for positively charged side chains. The catalytic triad comprises Ser195, His57 and Asp102. The joint backbone amides of Gly193 and Ser195 create the oxyanion hole (Figure 4.2 A).

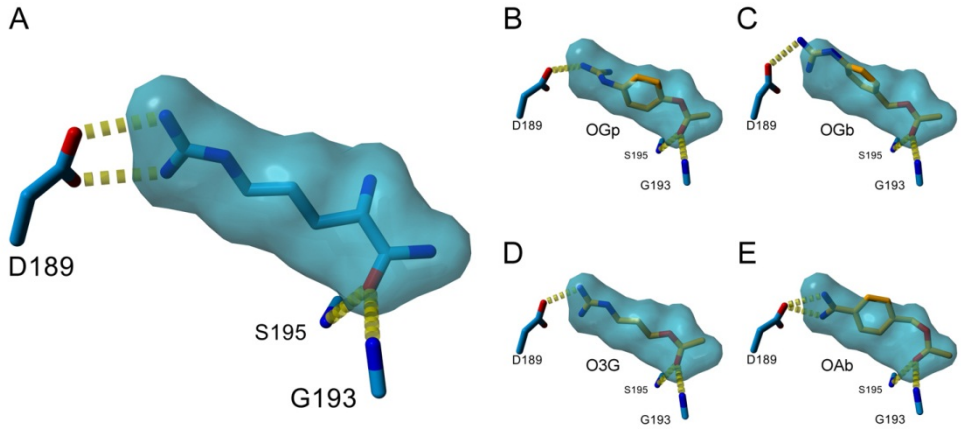


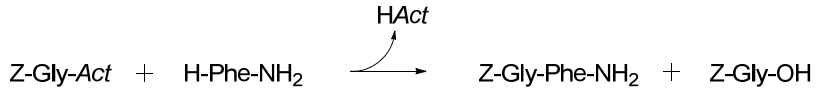
Figure 4.2 *Molecular modelling of OGp analogues in trypsin*

The Z-Gly-OGp ester (Figure 4.2 B) could be easily docked in the active site of trypsin, positioned in a similar way as the arginine side chain (Figure 4.2 A), which was in agreement with previous results from literature.[7] The guanidino group made the aforementioned crucial interaction with Asp189 in the S_1 pocket and, additionally formed hydrogen-bonds with Ser190 and Trp215. Furthermore, the substrate carbonyl was nicely located in the oxyanion hole. The docking poses of the benzylic esters Z-Gly-OGb and Z-Gly-OAb were both suboptimal, even though the carbonyl groups were located in the oxyanion hole. Figure 4.2 C shows that OGb slightly extended beyond the volume occupied by the arginine side chain as a result of the additional carbon atom, whereas OAb (Figure 4.2 E) has to adopt a somewhat distorted conformation to be located in the oxyanion hole while simultaneously retaining interaction between the amidinium group and Asp189. The aliphatic Z-Gly-O3G analogue (Figure 4.2 D) perfectly mimicked the side chain of arginine, thus achieving hydrogen bond interactions with Asp189, Ser190 and Tyr217. The oxyanion hole residues Gly193 and Ser195 were in the correct position to stabilise the carbonyl of the ester. Z-Gly-NGp showed a binding mode similar to Z-Gly-OGp.

4.2.2 Experimental activity of OGp analogues

Subsequently, the OGp analogues were evaluated experimentally in an enzymatic assay with trypsin. H-Phe-NH_2 was used as the acyl acceptor because of its distinctive UV properties at 254 nm, which simplifies HPLC analyses. The enzymatic reaction was monitored for three hours. The identity of the products was confirmed by chemical synthesis of reference compounds and LC-MS analysis. Table 4.1 presents either the time to reach full conversion, or the extent of conversion after three hours. The background hydrolysis of the analogues was determined from a blank reaction where no trypsin was present. The indicated percentages of enzymatic synthesis and hydrolysis remained constant over time, as measured after 24 hours, unless stated otherwise.

Table 4.1 Various Z-Gly-*Act* compounds tested experimentally[a]



				background	*enzymatic*	
Entry	**Act**	**Time (min)**	**Conv. (%)**	**Z-Gly-OH (%)**	**Z-Gly-Phe-NH_2 (%)**	**Z-Gly-OH (%)**
1	OGp	15	100	2.7	22.5	74.8
2	OAb	180	43	2.0	23.4	17.6
3	OGb	120	99	3.3	72.5	23.2
4	NGp	180	-	-	-	-
5	O3G	180	38	0.9	29.6	7.5
6[b]	O3G∇	180	73	5.0	54.0	14.0
7	O3G=	180	100	2.4	76.3	21.3
8	OTfe	90	100	5.7	73.6	20.7

[a] Conditions: 2 mM Z-Gly-*Act*, 15 mM H-Phe-NH_2, 0.2 M HEPES buffer pH 8.0, 0.2 M NaCl, 20 mM $CaCl_2$, 10% (v/v) DMF

[b] The entries below the line will be discussed in § 4.2.3

Z-Gly-OGp is readily converted by trypsin, although the synthesis over hydrolysis ratio is not too favourable (entry 1). The remaining analogues do not show any activity at the same enzyme concentration, except for Z-Gly-OGb, which was slightly active. A hundred-fold increase in trypsin concentration gave some differentiation: Z-Gly-OGb (entry 4) is almost completely consumed in two hours, whereas Z-Gly-OAb (entry 3) and Z-Gly-O3G (entry 6) are converted for 43 and 38%, respectively, in three hours. The lower activity of Z-Gly-OAb may be explained by its somewhat distorted fit in the active site of trypsin. The low activity of Z-Gly-O3G, however, is absolutely unexpected, as this analogue appeared to be a perfect mimetic. Z-Gly-NGp is the only analogue that remained completely inactive under these conditions. In all cases, the leaving group character of the analogues is reduced compared to OGp. Hence, we conclude that although all the analogues are expected to be fairly similar in terms of affinity for the enzyme, this contribution to activity is rather small. Inversely, these results imply a large influence of the leaving group ability of the ester.

4.2.3 O3G variants with increased activity

If electronic properties indeed play an important role, it should be possible to increase the activity of the analogues by transforming them into better leaving groups. For this purpose O3G was selected, as it appears to be a perfect mimic of the natural substrate and its activity is surprisingly low compared to OGp.

Inspection of the molecular model of trypsin with Z-Gly-O3G revealed that there is some space in the binding pocket for the introduction of a small substituent (Figure 4.3), such as

a methylene (O3G=) or two fluorides (O3GF_2), which are weak and strong inductively electron-withdrawing groups, respectively. However, these modifications also create additional Van der Waals interactions with the pocket, which by itself can be a reason for increased affinity of the ester for the enzyme. To assess this effect, the cyclopropyl (O3G▽) group was included, which was shown to fit in the binding pocket to make these additional interactions, without altering the electronic properties.

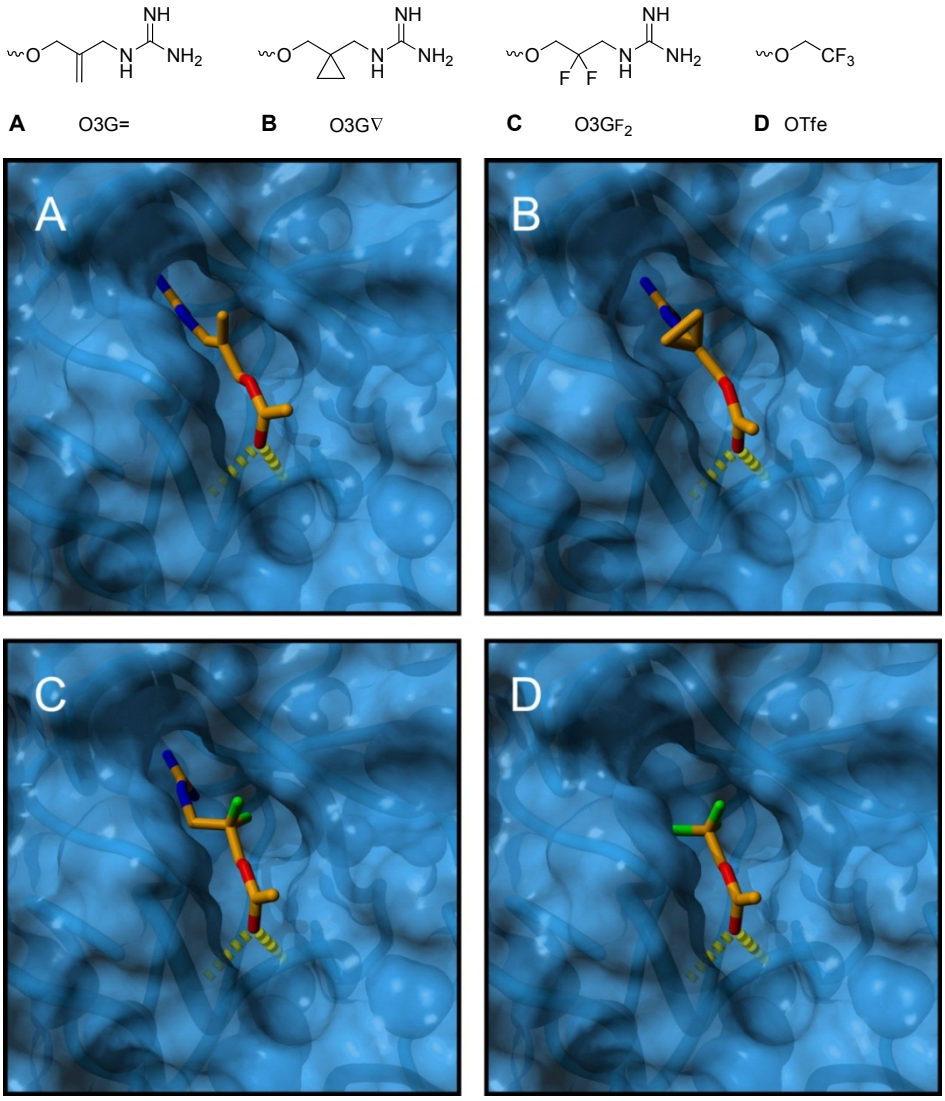


Figure 4.3 *Structures of O3G variants and their docking poses*

After the synthesis of the O3G variants, they were experimentally evaluated in the enzymatic assay with trypsin (Table 4.1). As the synthesis of Z-Gly-O3GF_2 failed (synthesis details are provided in Chapter 7), the trifluoroethyl ester (OTfe) was taken as an alternative, as it also nicely fits in the active site of trypsin (Figure 4.3 D) and it is known to act as a good leaving group. Both 03G variants showed considerable superior activity over O3G. As anticipated, the addition of the steric cyclopropyl group (entry 7) increased the activity (from 38 to 73% conversion in 180 min), whereas the introduction of the slightly inductively electron withdrawing methylene (entry 8) improved the activity even further (100% conversion in 180 min). Surprisingly, of all the analogues tested, Z-Gly-OTfe (entry 9), yielded the fastest reaction (100% conversion in 90 min), despite the fact that no cationic recognition element was present in the molecule. This again supports the idea that the leaving group ability is the main contributor to the suitability of the ester for enzymatic peptide synthesis.

4.2.4 *Ab initio* calculations

An appropriate computational technique to study electronic events within or between molecules is *ab initio* calculations. As the computational requirements are large for these calculations, only four compounds were selected, *i.e.* Z-Gly-OGp, Z-Gly-O3G, Z-Gly-NGp and Z-Gly-OTfe. Furthermore, the system was drastically simplified, in that the enzyme was represented by only the hydroxyl connected to a carbon of the active site serine. In addition, part of the protecting group of the ester was not taken into consideration. This situation is displayed in the white spot of Figure 4.4 A. The first step of the reaction comprises the formation of a tetrahedral intermediate, which is stabilised by the oxyanion hole (Figure 4.4 B). The subsequent collapse of the intermediate will liberate the alcohol, or amine in the case of NGp, from the complex (Figure 4.4 C). This is the endpoint of our calculations, as we are interested in differences in leaving group ability. Obviously, the enzyme still needs to be deacylated by a nucleophile to complete the catalytic cycle, which will also proceed through a tetrahedral intermediate.

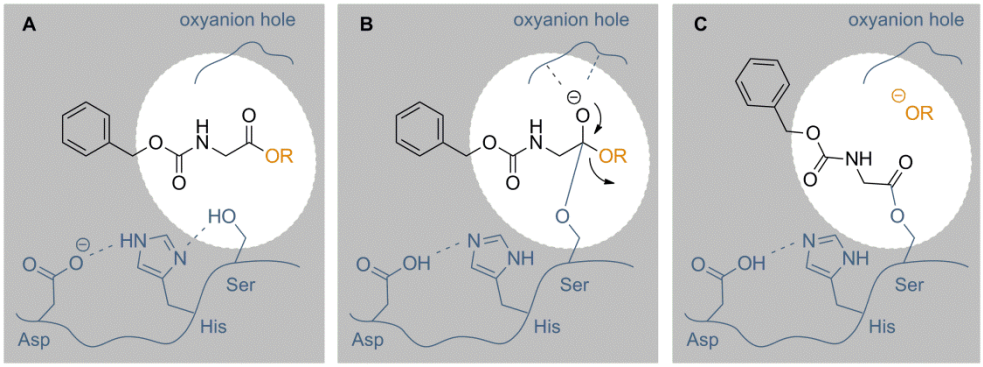


Figure 4.4 *Reaction path of Z-Gly-Act in trypsin*

We started out to locate the tetrahedral intermediate for the OGp compound. After exhaustive partial optimisations, in which the C-O bond connecting OGp to the central carbon of the tetrahedral intermediate was kept fixed, we came to the conclusion that nowhere along this internal coordinate did a stationary point exist. No optimum or saddle point could be found. Removing the restraint of the fixed C-O bond always resulted in dissociation upon optimisation. A stable intermediate was found, however, when a hydrogen atom was added to the oxygen carrying the negative charge. This is in line with the commonly accepted hypothesis[12] that, in the natural protein environment, hydrogen bonding in the oxyanion hole stabilises the tetrahedral intermediates. As an equivalent amount of atoms and charge is required throughout the calculations, a hydrogen atom was added to the leaving group too. This corresponds with the accepted mechanism that the leaving group is protonated upon formation of the first tetrahedral intermediate.[13] With the stable intermediates identified, the energy diagram as shown in Figure 4.5 and the accompanying Table 4.2 were produced.

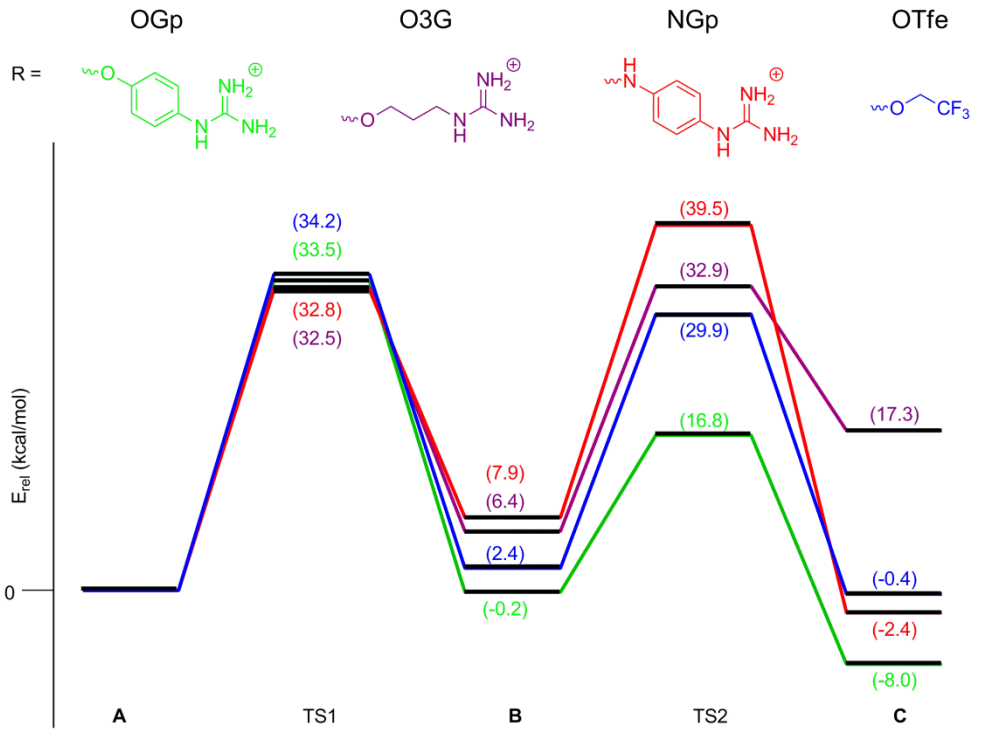
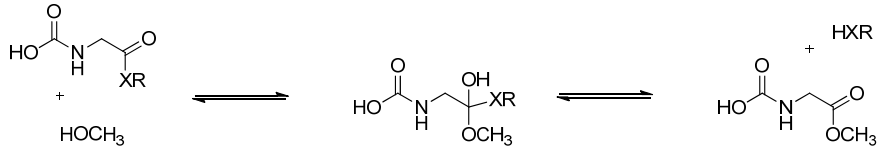


Figure 4.5 *Energy diagram derived from* ab initio *calculations. **A**, **B** and **C** correspond with the structures depicted in Table 4.2.*

Table 4.2 *Ab initio* computed energies



A TS1 **B** TS2 **C**

Compound (XR)	Reaction coordinate	E-B3LYP (Hartree)	ZPVE (Hartree)	E-relative (Kcal/mol)
OGp	A	-1023.822519	0.318313	0
	TS (A→B)	-1023.768428	0.317477	33.5
	B	-1023.827721	0.323082	-0.2
	TS(B→C)	-1023.796406	0.318871	16.8
	C	-1023.836929	0.319910	-8.0
O3G	A	-910.799298	0.326420	0
	TS (A→B)	-910.745545	0.324387	32.5
	B	-910.793383	0.330697	6.4
	TS(B→C)	-910.744129	0.323695	32.9
	C	-910.770669	0.325304	17.3
NGp	A	-1003.978561	0.333122	0
	TS (A→B)	-1003.923725	0.330513	32.8
	B	-1003.970149	0.337313	7.9
	TS(B→C)	-1003.913334	0.330891	39.5
	C	-1003.981686	0.332390	-2.4
OTfe	A	-964.639842	0.194069	0
	TS (A→B)	-964.582210	0.193946	34.2
	B	-964.638998	0.200064	2.4
	TS(B→C)	-964.588930	0.193792	29.9
	C	-964.639257	0.195890	-0.4

In the diagram, the relative energies are compared, so for convenience the position of the starting point for each of the reaction paths is arbitrarily set to zero. The activation energy (TS1) of the formation of tetrahedral intermediate (B) is quite similar for all compounds under investigation, although the energy level of B differs for each analogue. The next transition state (TS2), which is most indicative for the leaving group ability of the various alcohols and amine, shows more variation but with a similar trend. The enzyme is cooperative in stabilising the negative charge that develops on the leaving group during the collapse of the tetrahedral intermediate, which was simulated by a proton. However, the differences between the leaving groups are mainly determined by the extent to which they can stabilise the developing negative charge. The low activation energy (TS2) of OGp

reflects its ability to mesomerically stabilise the negative charge, whereas the high activation energy of NGp, also capable of mesomeric stabilisation, can be explained by the lower electronegativity of nitrogen compared to oxygen. Similarly, but to a lesser extent, the inductively electron-withdrawing effect of the OTfe group is helpful in stabilising the negative charge. In the case of O3G, the negative charge is isolated on the oxygen, without possibilities for further stabilisation, which may also account for the high energy level of situation C.

Upon linking these computational results to the experimental outcomes, one should bear in mind that only the acylation step was computationally studied. The experiments with trypsin provide insight in the efficiency of the complete catalytic cycle, including deacylation of the enzyme. In spite of this, the ranking OGp > OTfe > O3G > NGp from the *ab initio* calculations is in agreement with the experimentally determined activities of Z-Gly-OGp and its analogues. In addition, the marked contrast between the highly energetic tetrahedral intermediate and TS2 of NGp and the favourable energies for OGp, seem to be consistent with the observation that acylation is the rate determining step in amide hydrolysis[14] as opposed to deacylation being rate limiting in OGp ester hydrolysis.[15] However, according to Menger *et al.* the reaction kinetics of proteases towards esters are highly dependent of the nature of the ester, as was demonstrated by comparing the *p*-nitrophenyl ester with the ethyl ester.[16] These authors argue that the '*p*-nitrophenyl ester syndrome' can be attributed to excellent electrophilic assistance to the departing entity. When reasoning by this analogy, it follows that the O3G ester does not stand a chance against OGp.

4.3 Conclusion

Various methods were employed to determine the properties to which the success of the OGp ester as substrate mimetic and enzyme-specific activating ester can be attributed. Although a direct experimental approach to determine solely the affinity of a substrate is not available, a computational docking study of closely related OGp analogues was insightful.

Z-Gly-OGb and Z-Gly-OAb showed that a worse fit in the active site, could be directly linked to a decrease in activity in the enzymatic assay. This was anticipated, as a good fit in the active site can be considered as an indication for affinity. To our complete surprise, however, it was also demonstrated that Z-Gly-O3G, although being a perfect analogue of the natural substrate arginine according to docking studies, is barely active compared to Z-Gly-OGp. Apparently, recognition of the ester by the enzyme is alone insufficient for activity, in contrast to what has been suggested for substrate mimetics in literature.

We hypothesised that a major contribution was provided by the leaving group ability of the ester, which was supported by *ab initio* calculations showing that OGp is a good leaving group whereas NGp and O3G are not. Furthermore, we were able to design improved O3G variants based on this hypothesis. Increasing activity was observed in the order O3G <

O3G∇ < O3G= < OTfe, the latter being the most convincing one because, although no cationic recognition element is present in this molecule, it is a good leaving group.

4.4 Acknowledgements

B. Bögels is gratefully ackowledged for his contribution to the experimental part of this chapter. B. Zarzycka and dr. S. B. Nabuurs are gratefully ackowledged for performing the docking studies. Dr G. Schaftenaar is gratefully acknowledged for performing the *ab initio* calculations and Dr. M. C. Feiters assisted with the interpretation of the *ab initio* results, which is highly appreciated.

4.5 Experimental Section

Molecular modelling of trypsin – OGp analogue complexes:

All described molecular docking studies were performed using the flexible docking programme Fleksy.[8, 17] The crystal structure of trypsin in complex with bovine pancreatic trypsin inhibitor (BPTI)[18], solved at 1.5 Å resolution, was used as the receptor structure (PDB entry 3FP6). The structure was prepared for docking by removing BPTI and all water molecules from the complex. Subsequently, hydrogen atoms were added to the structure and their positions were optimised using the Yasara program[9]. In the applied docking protocol only those docking poses were taken forward in which the scissile bond of the docked substrate mimetic aligned to the scissile bond of the natural peptide substrate. Otherwise, default parameters as described previously[8] were applied.

General procedure for the enzymatic reactions:

Enzymatic acyl transfer reactions were performed at 25 °C in a total volume of 375 μL containing 0.2 M HEPES buffer (pH 8.0), 0.2 M NaCl, 20 mM $CaCl_2$, 10% DMF and 2 mM *p*TSA as an internal standard. Stock solutions of Z-Gly-*Act* compounds (50 mM) in DMF and H-Phe-NH_2 (30 mM) in buffer were prepared. The final concentrations of acyl donor and acyl acceptor were 2 mM and 15 mM, respectively. The latter was calculated as free, N^{α}-unprotonated nucleophile concentration $[HN]_0$ according to the Henderson–Hasselbalch equation $[HN]_0 = [N]_0/(1 + 10^{pK-pH})$. To trypsin (9.6 mg) was added milliQ (1 mL), the solution was stirred and stored for maximally one month in aliquots at –20 °C. Following thermal equilibration of assay mixtures, the enzymatic reactions were started by addition of trypsin at a final concentration of 1.6 or 160 μM. Blanc reactions were run in parallel, but milliQ was added instead of trypsin. From this control experiment the spontaneous ester hydrolysis could be determined, as well as non-enzymatic aminolysis of the acyl donor esters of which the latter could be ruled out. With regular intervals 20 μl aliquots were withdrawn and quenched with 20 μl glacial acetic acid. The reactions were monitored for 3 hours by HPLC and checked once more for changes in reaction mixture composition after 24 hours. The values reported are the average of at least two separate experiments. The identity of the formed peptide products was established by chemical synthesis of reference compounds and LC-MS.

HPLC-Analyses:

Samples were analyzed by a Shimadzu LC 2010 analytical HPLC system equipped with a RP C18 column (Varian, Inertsil ODS-3, 5 μm, 150 × 4.6 mm) and eluted with various mixtures of acetonitrile/water containing 0.1% trifluoroacetic acid under isocratic and gradient conditions at flow rates of 1.0 mL·min^{-1}. The wavelength of detection was 254 nm. Product yields were calculated from peak areas of the substrate esters and the hydrolysis and aminolysis products.

***Ab initio* calculations:**

Standard LCAO-MO-SCF calculations were performed with the programme Gamess-US[19] employing restricted Hartree-Fock (RHF) procedures and Density Functional Theory DFT). All DFT calculations were done using B3LYP exchange-correlation functional.[20] The geometries of the isomers were determined using analytical gradient and numerical second-derivative optimisation procedures with the 6-31G** basis set. The relative energies were corrected for the contribution of zero-point vibrational energies (ZPVE). The ZPVEs were calculated for the 6-31G** optimised geometries employing RHF and 6-31G** basis set.

4.6 References

[1] M. T. Reetz, *Angew. Chem., Int. Ed.* **2011**, *50*, 138-174.
[2] F. Bordusa, *Chem. Rev.* **2002**, *102*, 4817-4867.
[3] F. Bordusa, *Curr. Protein Pept. Sci.* **2002**, *3*, 159-180.
[4] R. J. A. C. de Beer, B. Zarzycka, H. I. V. Amatdjais-Groenen, S. C. B. Jans, T. Nuijens, P. J. L. M. Quaedflieg, F. L. van Delft, S. B. Nabuurs, F. P. J. T. Rutjes, *ChemBioChem* **2011**, *12*, 2201-2207.
[5] R. J. A. C. de Beer, B. Zarzycka, M. Mariman, H. I. V. Amatdjais-Groenen, M. J. Mulders, P. J. L. M. Quaedflieg, F. L. van Delft, S. B. Nabuurs, F. P. J. T. Rutjes, *ChemBioChem* **2012**, *13*, 1319-1326.
[6] (a) L. Polgár, *Cell. Mol. Life Sci.* **2005**, *62*, 2161-2172; (b) H. P. Kasserra, K. J. Laidler, *Can. J. Chem.* **1969**, *47*, 4031-4039.
[7] M. Thormann, S. Thust, W. H. Hofmann, F. Bordusa, *Biochemistry* **1999**, *38*, 6056-6062.
[8] S. B. Nabuurs, M. Wagener, J. de Vlieg, *J. Med. Chem.* **2007**, *50*, 6507–6518.
[9] E. Krieger, G. Koraimann, G. Vriend, *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 393-402.
[10] R. Huber, W. Bode, *Acc. Chem. Res.* **1978**, *11*, 114-122.
[11] I. Schechter, A. Berger, *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157-162.
[12] R. C. Wilmouth, K. Edman, R. Neutze, P. A. Wright, I. J. Clifton, T. R. Scheider, C. J. Schofield, J. Hajdu, *Nat. Struct. Biol.* **2001**, *8*, 689-694.
[13] B. Liu, C. J. Schofield, R. C. Wilmouth, *J. Biol. Chem.* **2006**, *281*, 24024-24035.
[14] D. M. Blow, *Acc. Chem. Res.* **1976**, *9*, 145-152.
[15] H. Sekizaki, K. Itoh, E. Toyota, K. Tanizawa, *Chem. Pharm. Bull.* **1996**, *44*, 1577-1579.
[16] F. M. Menger, M. Ladika, *J. Am. Chem. Soc.* **1987**, *109*, 3145-3146.
[17] M. Wagener, J. de Vlieg, S. B. Nabuurs, *J. Comput. Chem.* **2012**, *in press*.
[18] E. Zakharova, M. P. Horvath, D. P. Goldenberg, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11034–11039.
[19] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery Jr., *J. Comput. Chem.* **1993**, *14*, 1347-1363.
[20] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648-5652.

Chapter 5

α-Chymotrypsin-catalysed peptide bond formation

The concept of enzyme-specific activation has so far only been investigated in combination with the cysteine protease papain. To study the somewhat broader scope of this phenomenon, our set of potentially activating esters was applied to the serine protease α-chymotrypsin and a comparison between both enzymes was made. The benzyl (OBn) and dimethylaminophenyl (ODmap) esters appeared to be papain-specific, whereas the guanidinobenzyl ester (OGb) gave the best results with α-chymotrypsin. For the latter ester the scope of a model dipeptide synthesis reaction under aqueous conditions was investigated by varying the amino acid donor. Computational docking analysis of the set of esters was performed to provide insight into the difference in reactivity of all the potentially activating esters.

5.1 Introduction

The protease papain was able to effect dipeptide synthesis in aqueous environment in moderate to excellent S/H ratios with the OGp ester as a recognition group. Instead of functioning as a substrate mimetic as we had anticipated, we showed by analysing a large set of esters and conducting computational docking studies that the reaction proceeds through an unprecedented phenomenon which we called enzyme-specific activation.[1] Based on this insight, we developed the simplified alternative activating esters ODmap and OBn.[2] These scientifically interesting and unexpected results made us wonder whether the concept of specific activation would be applicable to other enzymes besides papain.

A preliminary screening with cheap and commercially available enzymes such as a protease from *Aspergillus oryzae,* bromelaine, pronase E and α-chymotrypsin, showed that results with the latter protease were most promising to pursue.

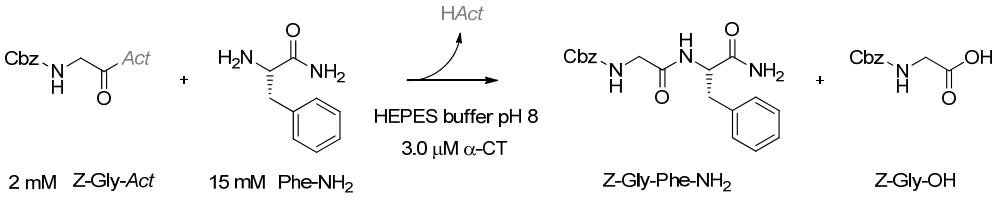
Chymotrypsin is a digestive enzyme produced in the pancreas as the inactive precursor chymotrypsinogen. Processing occurs with the help of trypsin and autolysis, resulting in varying amounts of α, π, δ and γ forms, of which the first is predominantly present.[3] This serine protease consists of 241 amino acid residues divided over three separate polypeptide chains (A, B and C) linked by disulfide bridges. α-Chymotrypsin selectively catalyses the hydrolysis of peptide bonds on the *C*-terminal side of tyrosine, phenylalanine, tryptophan and leucine.

5.2 Results and Discussion

5.2.1 Evaluation of potentially enzyme-specific activating esters

The set of potentially activating esters that was originally designed for papain (Figure 3.1) was not adapted to better match the specificity of α-chymotrypsin, since several esters already contained an aromatic part. It is described in literature that the OGp ester functions as a substrate mimetic for α-chymotrypsin,[4] although the terminology in this case is questionable, since the natural substrates of this protease do not contain a positively charged guanidino moiety. In addition, OTfe[5] and OCam[6] esters are known to broaden the substrate scope of α-chymotrypsin considerably.

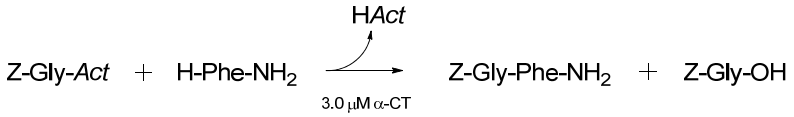
The esters were evaluated for their ability to form the dipeptide Z-Gly-Phe-NH_2. The conversions were determined using a previously described assay, but now using α-chymotrypsin instead of papain (Scheme 5.1).[1] The amount of α-chymotrypsin was adjusted such, that the activity of both enzymes with the Z-Gly-OGp ester was comparable.



Scheme 5.1 *Enzymatic dipeptide synthesis with α-chymotrypsin*

Table 5.1 shows either the time to reach 100% conversion, or the conversion after 3 hours. The background hydrolysis of the esters was determined from a blank reaction where no α-chymotrypsin was present. The indicated percentages of enzymatic synthesis and hydrolysis remained constant over time as measured after 24 hours.

Table 5.1 Various Z-Gly-*Act* esters tested experimentally[a]



				background	*enzymatic*	
Entry	***Act*[b]**	**Time (min)**	**Conv. (%)**	**Z-Gly-OH (%)**	**Z-Gly-Phe-NH_2 (%)**	**Z-Gly-OH (%)**
1	OGp	25	100	6.5	78.4	15.1
2	OGb	180	97	0.7	86.0	10.3
3	NGp	180	-	-	-	-
4	OAb	120	100	3.0	76.0	21.0
5	OTmap	15	100	13.1	74.4	12.5
6	ODmap	120	100	3.1	80.2	16.7
7	OTmape	180	50	1.9	35.5	12.6
8	ODmape	180	<1	0.6	0.2	0.2
9	O3G	180	32	2.3	22.4	7.3
10	O4G	180	51	2.0	36.1	12.9
11	O5G	180	89	2.4	63.8	22.8
12	O4A	180	13	0.3	9.4	3.3
13	O5A	180	32	1.5	23.6	6.9
14	OCam	120	100	9.2	73.9	16.9
15	O3Cam	180	9	0.7	7.4	0.9
16	O4Cam	180	20	1.0	14.6	4.4
17	OBn	180	80	0.3	57.3	22.4
18	ONb	180	31	[c]	10.4	20.6[c]
19	OTfe	45	100	3.5	74.5	22.0
20	OMe	180	7	0.6	4.9	1.5

[a] Conditions: 2 mM Z-Gly-*Act*, 15 mM H-Phe-NH_2, 0.2 M HEPES buffer pH 8.0, 0.2 M NaCl, 20 mM $CaCl_2$, 10% (v/v) DMF

[b] Structures can be found on the inside of the back cover flap

[c] Because the HOBn peak showed overlap with the Z-Gly-OH peak in the HPLC chromatogram, it was not possible to distinguish between spontaneous and enzymatic hydrolysis.

The fastest esters are OTmap (entry 5), OGp (entry 1) and OTfe (entry 19), which show an increasing amount of enzymatic hydrolysis in this order and a decreasing, but relatively high amount of background hydrolysis. The fact that OGp and OTfe are both active is in agreement with the literature.[4-5] The next fastest cluster of esters consists of OAb (entry 4), ODmap (entry 6) and OCam (entry 14). The S/H ratios are mutually comparable and in the same range as OTfe. The background hydrolysis of OCam is significantly higher than that of the other two. Miyazawa reported that the OCam ester is superior to the OTfe ester, which is not confirmed by our data. This deviation may be due to the different nature of the solvents, aqueous in our case, but nearly anhydrous for the case reported.[5-6]
Although the previously mentioned esters gave faster reactions, the best properties are actually found for the OGb ester (entry 2). The background hydrolysis is low, and it shows the best S/H ratio, which are both desirable properties. At this point, it is noteworthy to pay some attention to the differences in performance of the various benzylic esters. The benzyl ester with no substituents (OBn, entry 17) reacts more slowly and in a worse S/H ratio than both OAb and OGb (entries 4 and 2, respectively). The addition of a nitro group on the para position (ONb, entry 18), which improves the leaving group ability, only deteriorates the situation. Apparently, a positively charged substituent is most favourable.
As could be expected from their electronic properties, OTmape, ODmape (entries 7 and 8, respectively) and most of the aliphatic esters are hardly active, with the exception of O5G (entry 11). Upon closer inspection, a trend becomes visible that comprises an increasing activity with increasing length of the aliphatic chain. This holds true for the sequence O5G>O4G>O3G as well as for O5A>O4A and O4Cam>O3Cam. Remarkable in this respect is that the largest effect is again accomplished by the positively charged guanidino group.

5.2.2 Activity in α-chymotrypsin versus papain

To properly estimate the enzyme-dependence in our set of potentially activating esters, the results from papain and α-chymotrypsin were compiled in one diagram for convenience of comparison (Figure 5.1).
The most noticeable difference between the two enzymes is that the percentage of enzymatic hydrolysis is structurally higher for all esters in the case of α-chymotrypsin. Moreover, the reactions generally proceed more slowly. The background hydrolysis is very well comparable, as it should be, since the blanc experiments are actually exact copies of each other. Only OCam shows a higher amount of background hydrolysis with α-chymotrypsin, which can be explained by the eight times longer reaction time that was required for complete conversion (15 versus 120 minutes).

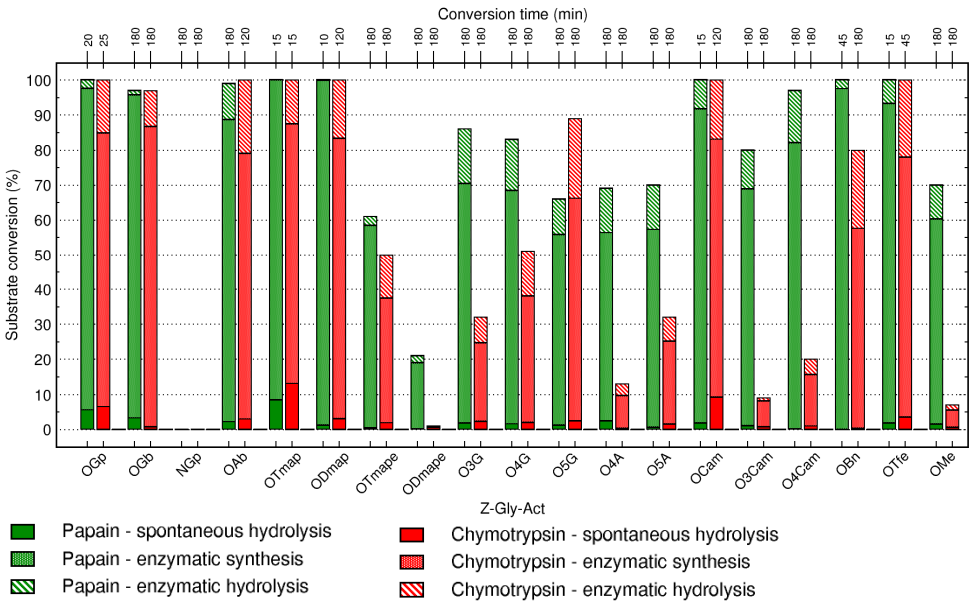


Figure 5.1 *Visual comparison of activating esters in papain and α-chymotrypsin. The diagram is composed of data from Tables 3.1 and 5.1*

The trend that was observed of higher reactivity with increasing aliphatic chain length in α-chymotrypsin, is less obvious in papain. The correlation seems to be reversed for the guanidino containing compounds (O3G, O4G and O5G), absent for the amino containing compounds (O4A and O5A) and less profound in the case of the compounds with a carbamoyl moiety present (O3Cam and O4Cam). Important to notice is that the OBn and ODmap esters indeed are specifically activating esters for papain, as they are not particularly activating with α-chymotrypsin.

5.2.3 Docking of activating esters into the active site of α-chymotrypsin

Molecular docking studies have been valuable before, so that this technique was also applied to α-chymotrypsin. The three dimensional structure of α-chymotrypsin has been previously solved by crystallographic studies.[7] The three interlinked peptide chains form an enzyme structure that is relatively similar to trypsin. Five subpockets (S_3 through S_2') are involved in the specific binding of the substrate, of which S_1 is the most important. Contrary to trypsin, where interactions with Asp189 are most determining for substrate specificity, the S_1 pocket of α-chymotrypsin does not contain one such crucial residue. In this protease, the pocket is characterised as a rather large, highly hydrophobic cavity with the ability to fit in moieties as big as an indole ring. Multiple contacts, especially involving Ser190, Cys191-220, Val213, Trp215 and Tyr228, contribute to the preference for large hydrophobic amino acids. Furthermore, α-chymotrypsin contains the same catalytic residues as trypsin, namely Ser195, His57 and Asp102. Likewise, the oxyanion hole is

created by the backbone amides of Gly193 and Ser195.[8] These important residues are all indicated in Figure 5.2.

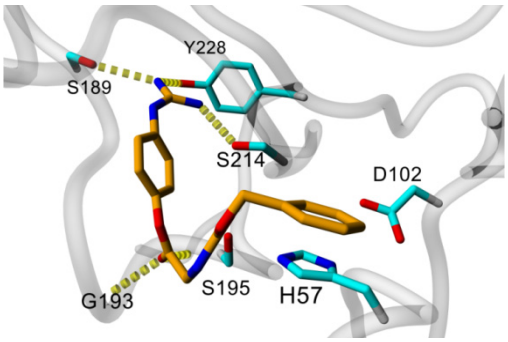


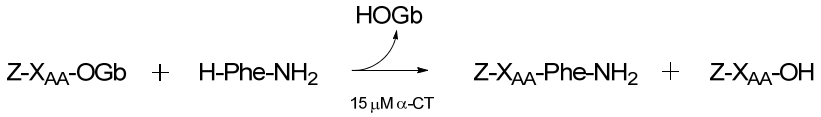
Figure 5.2 *Model of Z-Gly-OGp in α-chymotrypsin*

The Z-Gly-OGp ester (Figure 5.2) showed a comparable binding mode as predicted by Günther *et al*.[9] The guanidino group is part of a hydrogen bonding network, which probably accounts for the good results that were obtained with esters containing a positively charged substituent. Instead of discussing the docking studies in detail, a summary is given in the sequel. The S_1 pocket of α-chymotrypsin is so large, that all our esters fit in, with the exception of ODmape and OTmape (Table 5.1, entries 7 and 8), which are too large. Several amino acid residues present in the cavity are available for hydrogen bonding. The differences in activity of the various esters can mostly be attributed to differences in their ability to actually make these hydrogen bonding interactions.

5.2.4 Scope of the amino acid donor of Z-X_{AA}-OGb

As OGb appeared to be the best alternative for OGp, with respect to the decreased background hydrolysis and increased S/H ratio, the scope of the acyl donor was experimentally determined using this ester. A representative set of amino acids was selected and the corresponding OGb esters were synthesised (see § 7.2.2). The amount of α-chymotrypsin in the enzymatic synthesis assay was increased five-fold to compensate for the slower reaction rate. The identity of the products in the enzymatic reaction was confirmed by chemical synthesis of reference compounds and LC-MS analysis. Table 5.2 presents either the time to reach full conversion of the OGb esters, or the conversion after three hours.

Table 5.2 Various Z-X_{AA}-OGb esters tested experimentally[a]



				background	*enzymatic*	
Entry	**Amino acid**	**Time (min)**	**Conv. (%)**	**Z-X_{AA}-OH (%)**	**Z-X_{AA}-Phe-NH_2 (%)**	**Z-X_{AA}-OH (%)**
1	L-Ala	45	100	2.6	54.4	43.0
2	Gly	45	100	2.4	86.0	11.6
3	L-Gln[b]	5	100	0.7	42.3	57.0
4	L-Ile	180	100	-	51.5[c]	49.5
5	L-Lys[b]	15	100	[d]	32.6	67.4[d]
6	L-Phe	0.5	100	-	70.1	29.9
7	D-Phe	90	100	-	70.8	29.2
8	L-Pro	180	-	-	-	-
9	D-Ala	180	36	-	11.3	24.7
10	β-Ala	180	37	-	6.7	30.3

[a] Conditions: 2 mM Z-X_{AA}-OGp, 15 mM H-Phe-NH_2, 0.2 M HEPES buffer pH 8.0, 0.2 M NaCl, 20 mM $CaCl_2$, 10% (v/v) DMF

[b] Measured once, molar extinction coefficients estimated based on similar compounds

[c] Dipeptide product precipitated during reaction, this is the estimated yield

[d] Because the Z-L-Lys-OGb peak showed overlap with the Z-L-Lys-OH peak in the HPLC chromatogram, it was not possible to distinguish between spontaneous and enzymatic hydrolysis.

The reaction times as well as the S/H ratios vary considerably for the different amino acids. Glycine (entry 2) reacts fast with the most synthesis and β-alanine (entry 10) reacting slowly with the worst ratio. Z-L-Phe-OGb (entry 6) reacts really fast, which can be explained by the fact that phenylalanine is also naturally recognised by α-chymotrypsin. Interestingly, mainly dipeptide product is formed instead of hydrolysis taking place. It is noticeable that Z-D-Phe-OGb (entry 7), its enantiomer, results in the same S/H ratio, albeit at a much slower rate. The composition of the reaction mixtures was checked after 24 hours, showing that over time, all the Z-X_{AA}-Phe-NH_2 products were partially enzymatically converted into Z-X_{AA}-Phe-OH. Non-enzymatic side product formation as found for some Z-X_{AA}-OGp esters, was not observed for any of the OGb esters, probably because OGb is a worse leaving group.

5.2.5 Future plans

Another important reason to select OGb for further evaluation was the expectation that this moiety is accepted by Alcalase-CLEA in organic solvent, so that the ester can be

synthesised enzymatically. However, a preliminary study in which diBoc protected HOGb was subjected to Alcalase-CLEA in organic solvent in the presence of an acyl donor, showed only a minor amount of the desired product. From the HPLC chromatogram was deduced that the alcohol was not stable and rapidly decomposed.

As became clear from a comparison of the results of papain and α-chymotrypsin, the latter shows more enzymatic hydrolysis. In spite of this, we hypothesise that the advantage of a large S_1 pocket can be exploited by converting the enzyme from a serine into a cysteine protease. This strategy to improve the S/H ratio has been successfully applied before to subtilisin.[10] Due to the complexity of the post-translational processing of chymotrypsinogen to a mature protease, chymotrypsin has not been obtained as a recombinant protein produced by microorganisms. This excludes the genetic introduction of a point mutation. However, a chemical way to modify the active site is available.[11]
To specifically target the serine in the active site, the serine protease inhibitor phenylmethanesulfonyl fluoride (PMSF) is employed (Figure 5.3). Because of its increased nucleophilic character, only this serine is assumed to react. The hydroxyl is in this way converted into a good leaving group, which can then be substituted by addition of thioacetate. Subsequent spontaneous hydrolysis will result in the desired cysteine protease.

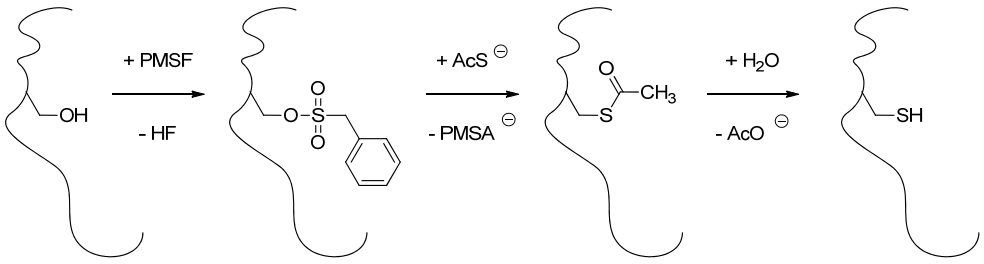


Figure 5.3 *Procedure for the chemical conversion of an active site serine into a cysteine*

A possible bottleneck in the application of this procedure to α-chymotrypsin is the presence of several disulfide bridges that hold together the three polypeptide chains. The excess of thioacetate used for the substitution reaction might cause denaturation and inactivity of the enzyme. However, preliminary treatment of α-chymotrypsin with an excess of thioacetate for 24 hours revealed that the activity with a test substrate was completely retained, which is an encouraging finding to start with. Further experiments on this procedure, however, were no further pursued.

5.3 Conclusion

Several esters of our set were successfully used in peptide synthesis catalysed by α-chymotrypsin, among which the OGp, OTmap, OCam and OTfe esters. The best result with

respect to S/H ratio was, however, obtained with the OGb ester. Subsequent docking studies showed that the S_1 pocket of α-chymotrypsin is rather large and hydrophobic, but in addition offers various possibilities for hydrogen bonding interactions, which may be an explanation for the apparent preference for guanidinium containing esters. It is notable that OBn and ODmap, which are activating esters for papain, do not specifically activate α-chymotrypsin. Thus, the activating properties of esters are indeed enzyme dependent.
Compared to papain, α-chymotrypsin displays a relatively high amount of enzymatic hydrolysis, even in the case of OGb. On the other hand, a pocket with ample space to accommodate many substrates might be worthwhile to further explore, provided that the S/H ratio can be improved.

5.4 Acknowledgements

H. I. V. Amatdjais-Groenen is gratefully ackowledged for her contribution to the experimental part of this chapter. B. Zarzycka and dr. S. B. Nabuurs are gratefully ackowledged for performing the computational studies.

5.5 Experimental section

General procedure for the enzymatic reactions:

Enzymatic acyl transfer reactions were performed at 25 °C in a total volume of 375 μL containing 0.2 M HEPES buffer (pH 8.0), 0.2 M NaCl, 20 mM $CaCl_2$, 10% DMF and 2 mM *p*TSA as an internal standard. Stock solutions of Z-Gly-*Act* or Z-X_{AA}-OGb compounds (50 mM) in DMF and H-Phe-NH_2 (30 mM) in buffer were prepared. The final concentrations of acyl donor and acyl acceptor were 2 mM and 15 mM, respectively. The latter was calculated as free, N^{α}-unprotonated nucleophile concentration $[HN]_0$ according to the Henderson–Hasselbalch equation $[HN]_0 = [N]_0/(1 + 10^{pK-pH})$. To chymotrypsin (18.75 mg) was added milliQ (1 mL), the solution was stirred and stored for maximally one month in aliquots at −20 °C. Following thermal equilibration of assay mixtures, the enzymatic reactions were started by addition of α-chymotrypsin at a final concentration of 3.0 or 15 μM. Blanc reactions were run in parallel, but milliQ was added instead of trypsin. From this control experiment the spontaneous ester hydrolysis could be determined, as well as non-enzymatic aminolysis of the acyl donor esters of which the latter could be ruled out. With regular intervals 20 μl aliquots were withdrawn and quenched with 20 μl glacial acetic acid. The reactions were monitored for 3 hours by HPLC and checked once more for changes in reaction mixture composition after 24 hours. The values reported are the average of at least two separate experiments. The identity of the formed peptide products was established by chemical synthesis of reference compounds and LC-MS.

HPLC-Analyses:

Samples were analyzed by a Shimadzu LC 2010 analytical HPLC system equipped with a RP C18 column (Varian, Inertsil ODS-3, 5 μm, 150 × 4.6 mm) and eluted with various mixtures of acetonitrile/water containing 0.1% trifluoroacetic acid under isocratic and gradient conditions at flow rates of 1.0 mL·min^{-1}. The wavelength of detection was 254 nm. Product yields were calculated from peak areas of the substrate esters and the hydrolysis and aminolysis products.

Molecular modelling of α-chymotrypsin – Z-Gly-*Act* complexes:

All described molecular docking studies were performed using the flexible docking programme Fleksy.[12] The crystal structure of α-chymotrypsin in complex with bovine pancreatic trypsin

inhibitor (BPTI)[13], solved at 1.7 Å resolution, was used as the receptor structure (PDB entry 1T8O). The structure was prepared for docking by removing BPTI and all water molecules from the complex. Subsequently, hydrogen atoms were added to the structure and their positions were optimised using the Yasara program[14]. In the applied docking protocol only those docking poses were taken forward in which the scissile bond of the docked substrate mimetic aligned to the scissile bond of the natural peptide substrate. Otherwise, default parameters as described previously[12a] were applied.

5.6 References

[1] R. J. A. C. de Beer, B. Zarzycka, H. I. V. Amatdjais-Groenen, S. C. B. Jans, T. Nuijens, P. J. L. M. Quaedflieg, F. L. van Delft, S. B. Nabuurs, F. P. J. T. Rutjes, *ChemBioChem* **2011**, *12*, 2201-2207.

[2] R. J. A. C. de Beer, B. Zarzycka, M. Mariman, H. I. V. Amatdjais-Groenen, M. J. Mulders, P. J. L. M. Quaedflieg, F. L. van Delft, S. B. Nabuurs, F. P. J. T. Rutjes, *ChemBioChem* **2012**, *13*, 1319-1326.

[3] S. T. Freer, J. Kraut, J. D. Robertus, H. T. Wright, N. H. Xuong, *Biochemistry* **1970**, *9*, 1997-2009.

[4] R. Günther, S. Thust, H. J. Hofmann, F. Bordusa, *Eur. J. Biochem.* **2000**, *267*, 3496-3501.

[5] T. Miyazawa, S. Nakajo, M. Nishikawa, K. Imagawa, R. Yanagihara, T. Yamada, *J. Chem. Soc. Perkin Trans. 1* **1996**, 2867-2868.

[6] T. Miyazawa, E. Ensatsu, N. Yabuuchi, R. Yanagihara, T. Yamada, *J. Chem. Soc. Perkin Trans. 1* **2002**, 390-395.

[7] (a) B. W. Matthews, P. B. Sigler, R. Henderson, D. M. Blow, *Nature* **1967**, *214*, 652-656; (b) H. Tsukada, D. M. Blow, *J. Mol. Biol.* **1985**, *184*, 703-711.

[8] D. M. Blow, *Acc. Chem. Res.* **1976**, *9*, 145-152.

[9] R. Günther, C. Elsner, S. Schmidt, H. J. Hoffman, F. Bordusa, *Org. Biomol. Chem.* **2004**, *2*, 1442-1446.

[10] T. Chang, D. Y. Jackson, J. P. Burnier, J. A. Wells, *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 12544-12548.

[11] (a) L. Polgár, M. L. Bender, *J. Am. Chem. Soc.* **1966**, *88*, 3153-3154; (b) K. E. Neet, D. E. J. Koshland, *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *56*, 1606-1611.

[12] (a) S. B. Nabuurs, M. Wagener, J. de Vlieg, *J. Med. Chem.* **2007**, *50*, 6507–6518; (b) M. Wagener, J. de Vlieg, S. B. Nabuurs, *J. Comput. Chem.* **2012**, *in press*.

[13] H. Czapinska, R. Helland, A. O. Smalås, J. Otlewski, *J. Mol. Biol.* **2004**, *344*, 1005-1020.

[14] E. Krieger, G. Koraimann, G. Vriend, *Proteins: Struc., Funct., Bioinf.* **2002**, *47*, 393-402.

Chapter 6

Alcalase-catalysed peptide bond formation

Chemoenzymatic dipeptide synthesis was investigated using the previously described set of esters in organic solvent with Alcalase-CLEA. The goal of this exercise was twofold: (1) study the performance of our activated esters in a chemoenzymatic system that was recently developed by DSM, and (2) investigate whether our set of esters could successfully be applied in dipeptide synthesis under anhydrous conditions.

Part of this chapter will be published:
R. J. A. C. de Beer,* T. Nuijens,* L. Wiermans, P. J. L. M. Quaedflieg, F. P. J. T. Rutjes, *Accepted in Org. Biomol. Chem.* **2012**, doi 10.1039/C2OB25662B.

* Both authors contributed equally

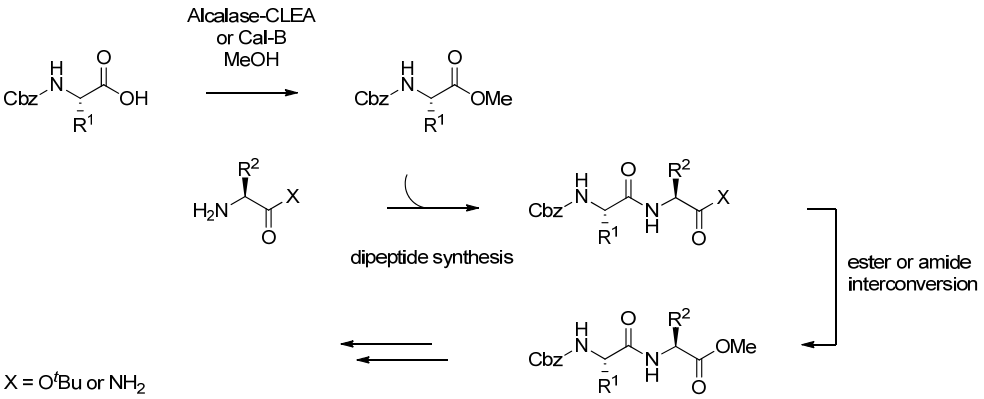
6.1 Introduction

Alcalase is the brand name for subtilisin A (Subtilisin Carlsberg), a serine endoprotease from *Bacillus licheniformis*.[1] Because this enzyme possesses broad substrate tolerance, with a preference for large hydrophobic residues in P_1, it is applied in the synthesis of numerous peptides.[2] Moreover, it is an inexpensive industrial protease (from Novozymes, 10 wt% solution, 6.75 €/kg).

In contrast to the dipeptide formation that was observed for papain, trypsin and α-chymotrypsin with our set of esters (Chapters III, IV and V, respectively), Alcalase only gave fast and complete hydrolysis of the starting material in all cases. Hence it was concluded that in order to explore the potential of this relatively unspecific enzyme for dipeptide formation, a system devoid of water was required.

Recently, Nuijens *et al*. noticed that Alcalase is capable of catalysing an esterification in *tert*-butanol at low water content to give *tert*-butyl esters from *N*-protected amino acids.[3] Normally, the water produced during this reaction and the water required for enzyme activity prohibits a favourable position for the esterification equilibrium resulting in low to moderate yields. However, upon the addition of molecular sieves, high yields were obtained since the continuous removal of water shifted the equilibrium towards product formation. Moreover, the enzyme apparently remained active under these conditions. Azeotropic distillation, which is much more efficient for large-scale reactions than molecular sieves, also efficiently removed the water. Further improvement of the system included the immobilisation of Alcalase as a CLEA,[4] so that the enzyme could easily be recovered by filtration and recycled with minimal loss of activity.

In addition to ester synthesis, the Alcalase system proved to be suitable for peptide synthesis as well. *C*-Terminal deprotection and activation was achieved in one single step by interconverting *tert*-butyl esters into primary alkyl esters, in particular methyl esters, which could then be coupled to the next amino acid (Scheme 6.1).[5] Besides *tert*-butyl esters α-carboxamides could be converted into primary alkyl esters in a similar process.[6]



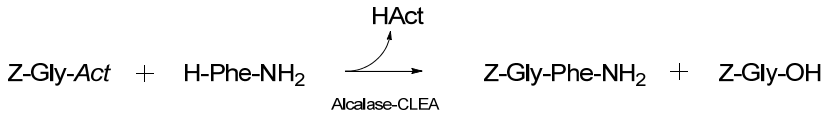
Scheme 6.1 *Chemoenzymatic peptide synthesis in the N→C direction using C-terminal t-Bu ester or carboxamide to ester interconversion.*

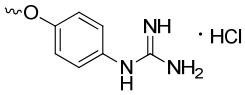
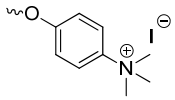
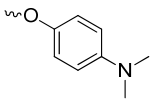
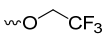
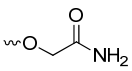
To broaden the scope of Alcalase-catalysed peptide synthesis even further, OCam and OTfe esters were introduced as acyl donors.[7] These moieties are known to be specifically recognised by some enzymes,[8] thereby overcoming their primary specificity. The coupling of challenging substrates such as sterically demanding amino acids as acyl donors (valine, isoleucine, threonine), notoriously weak nucleophiles (proline), or D- and other non-proteinogenic amino acid residues, became well feasible in high yields without any hydrolytic side reactions. However, coupling reactions sometimes remained rather slow and a relatively large amount of enzyme was required, especially when longer peptide fragments were to be coupled. Clearly, there is room for improvement of the acyl donor ester in Alcalase catalysed peptide synthesis.

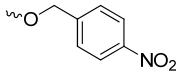
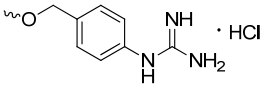
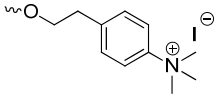
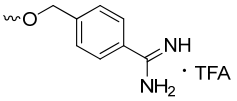
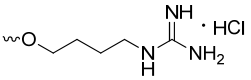
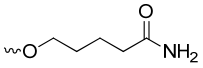
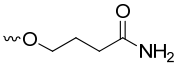
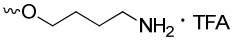
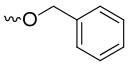
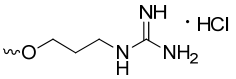
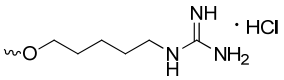
6.2 Results & Discussion

The available library of Z-Gly-*Act* esters was employed to test the coupling efficiencies in Alcalase-catalysed dipeptide synthesis, with H-Phe-NH_2 functioning as the nucleophile. A solvent mixture of DMF/THF (1:9 (v/v)) ensured good solubility of all starting materials. Molecular sieves were added to remove water from the reaction mixture, preventing any hydrolytic side reactions and Alcalase-CLEA facilitated convenient handling and workup. The results of the screening are given in Table 6.1.

Table 6.1 Relative activity of various Z-Gly-*Act* esters[a]



Entry	*Act*	Structure	Relative activity (%)
1	OGp		100
2	OTmap		95
3	ODmap		70
4	OTfe		68
5	OCam		60

6	ONb		34
7	OGb		31
8	OTmape		30
9	OAb		28
10	O4G		23
11	O4Cam		11
12	O3Cam		9
13	O4A		8
14	OBn		8
15	O3G		7
16	O5G		4
17	O5A		4

[a] Conditions: 2 mM Z-Gly-*Act*, 0.33 M H-Phe-NH_2, DMF/THF, crushed 3Å mol sieves, 50 °C, 60 min

Clearly, the substituted phenyl esters (entries 1-3) are the most active species for the Alcalase-CLEA catalysed peptide coupling. Especially for entries 1 and 2 this is remarkable, since Alcalase has a preference for large uncharged hydrophobic residues in both the P_1 and the P_1' positions.[9] The OTfe and OCam ester derivatives follow closely (entries 4 and 5). It is believed that the amide group of the OCam ester moiety (Figure 6.1 B) binds to the enzyme via a hydrogen bond in the same fashion as an amide of a peptide backbone (Figure 6.1 A) binds when it is recognised and cleaved by an endoprotease.

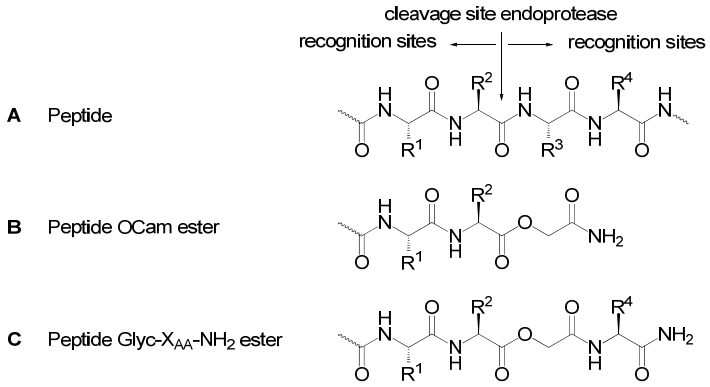


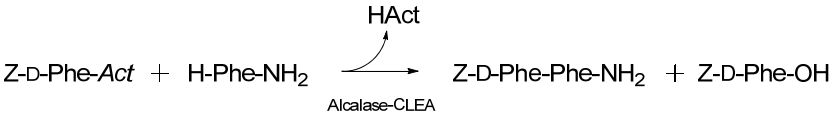
Figure 6.1 *Similarity of the natural peptide at the cleavage site of an endoprotease and the OCam and OCam-X_{AA}-NH_2 ester.*

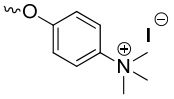
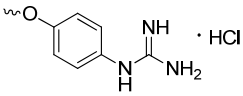
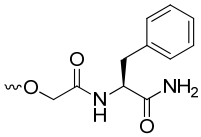
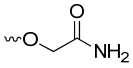
The relative activities of the substituted benzyl (entries 6, 7 and 9) and phenethyl (entry 8) esters are mutually comparable, yet significantly decreased compared to entries 1-5. An exception is the benzyl ester (entry 14), which clusters with the aliphatic esters (entries 11-13 and 15-17) in the low activity range. This contrasts with the high activity of the benzyl ester in combination with the enzyme papain.[10] Apparently, the length of the carbon linker is important, as demonstrated by the O4G ester (entry 10), which clearly outperformed both the O3G and O5G ester (entries 15 and 16, respectively). Generally, it is noticeable that high activity seems to coincide with strong electron-withdrawing properties of the esters, which render the carbonyl more susceptible to nucleophilic attack by the active site serine of the protease and the corresponding alcohol a better leaving group.

An important drawback of phenyl esters is that, due to their high activation level, they are relatively difficult to synthesise, while both the chemical[11] and enzymatic[7] synthesis of the OCam ester (entry 5) is well feasible. Analogous to the strategy followed by Wells *et al.*, the OCam ester can presumably be improved by elongating it with an amino acid amide, thereby creating additional binding interactions with the enzyme (Figure 6.1 C).[12] To this end, a library of Fmoc-Val-Ala-OGlyc-X_{AA}-NH_2 esters was evaluated, where X_{AA} stands for all 20 proteinogenic amino acids, with either a protected or an unprotected side chain functionality. The best results were obtained with the OGlyc-Phe-NH_2 ester, resulting in a two-fold enhancement.[13]

This optimal OGlyc-Phe-NH_2 ester was compared to the most active substituted phenol esters from the first Z-Gly-*Act* screening (Table 6.1, entries 1, 2). To clearly discern the intrinsic reactivities, the challenging substrate Z-D-Phe-OH was chosen as the acyl donor with H-Phe-NH_2 as the nucleophile (Table 6.2). D-Amino acids are notoriously difficult substrates for Alcalase, in fact, Chen *et al.* reported that no peptide product was obtained at all using Alcalase and Z-D-Phe as acyl donor.[2]

Table 6.2 Relative activity of various Z-D-Phe-*Act* esters[a]



Entry	*Act*	Structure	Relative activity (%)
1	OTmap		100
2	OGp		98
3	OGlyc-Phe-NH_2		90
4	OCam		54

[a] Conditions: 2 mM Z-Gly-*Act*, 0.33 M H-Phe-NH_2, DMF/THF, crushed 3Å mol sieves, 50 °C, 60 min

As is evident from Table 6.2, the elongated OCam ester (entry 3) shows a comparable reactivity as the substituted phenyl esters (entry 1, 2). An equally active but more conveniently accessible ester was thus developed for Alcalase-CLEA catalysed peptide synthesis. Another advantage is that no racemisation occurred on the activated amino acid ester, *i.e.* D-Phe, using the Z-D-Phe-OGlyc-Phe-NH_2 ester (entry 3, e.e. of D-Phe >99.5), this is in contrast with the Cbz-D-Phe-OTmap ester (entry 1, e.e. of D-Phe 85.8%).

6.3 Conclusion

Summarising, our set of activated esters was subjected to the enzyme Alcalase in anhydrous organic solvents. With this system developed by DSM, it became possible to shift the Alcalase activity entirely from hydrolysis in aqueous buffers to synthesis in the absence of water. To our surprise, it appeared that positively charged phenolic esters were most active. It was also demonstrated that the activity of OCam esters can be increased to the level of the most active phenolic esters by elongation with (apolar) amino acids and amino acid amides.

6.4 Acknowledgements

T. Nuijens, L. Wiermans and dr. P. J. L. M. Quaedflieg (DSM Innovative Synthesis B.V.) are kindly acknowledged for the fruitful collaboration.

6.5 Experimental Section

General remarks:

Before use, 3 g Alcalase-CLEA (Type OM, CLEA-Technologies, 580 U/g) was suspended in 100 mL ^{t}BuOH and crushed with a spatula. After filtration, the enzyme was resuspended in 50 mL MTBE followed by filtration. Large enzyme particles were removed by a sieve (0.5 mm pore size). Analytical HPLC chromatograms were recorded on an HP1090 Liquid Chromatograph, using a reversed-phase column (Phenomenex, C18, 5 µm particle size, 150 × 4.6 mm) at 40°C. The gradient program was: 0-25 min linear gradient ramp from 5% to 98% eluent B and from 25.1-30 min with 5% eluent B (eluent A: 0.5 mL/L methane sulfonic acid (MSA) in H_2O, eluent B 0.5 mL/L MSA in acetonitrile). The flow was 1 mL/min from 0-25.1 min and 2 mL/min from 25.2-29.8 min, then back to 1 mL/min until stop at 30 min. Injection volumes were 20 µL. The 3 Å molecular sieves (Acros, 8 to 12 mesh) were activated (200°C under vacuum overnight), crushed and sieved (0.5 mm pore size) to remove large particles. To determine the e.e. of Phe the samples concentrated *in vacuo* and the residue suspended in excess 6 N HCl and refluxed overnight. Chiral HPLC was performed on a crownether (+) column (150 mm length, 4.0 mm internal diameter, 5 µm particle size) at 25 °C with 30 mM aqueous $HClO_4$ (pH = 2.0) as the eluent. UV detection was performed at 210 nm using a UV-VIS linear spectrometer. The flow was 1 mL/min. Injection volumes were 5 µL. R_t (D-Phe) = 6.90 min, R_t (L-Phe) = 8.82 min.

General procedure for the relative activity determination of Alcalase-CLEA catalysed peptide coupling in organic media:

To a suspension of Alcalase-CLEA (4.5 mg), H-Phe-NH_2 (0.54 mg) and crushed 3 Å molecular sieves (4.5 mL) in THF (900 µL), was added amino acid or peptide ester stock solution in DMF (20 mM, 100 µL). The reaction mixture was shaken at 50 °C with 200 rpm for 60 min. Afterwards, the reaction mixtures were filtrated and analysed by analytical HPLC by integrating the peptide coupling product peak. Integration areas of different reactions were compared to determine the relative activity (ester which gave the highest peptide product intergration area = 100%).

6.6 References

[1] (a) S.-T. Chen, K.-T. Wang, C.-H. Wong, *J. Chem. Soc., Chem. Commun.* **1986**, *20*, 1514-1517; (b) R. J. Siezen, J. A. M. Leunissen, *Protein Sci.* **1997**, *6*, 501-523; (c) E. I. Smith, R. J. DeLange, W. H. Evans, M. Landon, F. S. Markland, *J. Biol. Chem.* **1968**, *243*, 2184-2191.

[2] S.-T. Chen, S.-Y. Chen, K.-T. Wang, *J. Org. Chem.* **1992**, *57*, 6960-6965.

[3] T. Nuijens, C. Cusan, J. A. W. Kruijtzer, D. T. S. Rijkers, R. M. J. Liskamp, P. J. L. M. Quaedflieg, *Synthesis* **2009**, *5*, 809–814.

[4] R. A. Sheldon, *Biochem. Soc. Trans.* **2007**, *35*, 1583-.

[5] T. Nuijens, C. Cusan, T. J. G. M. van Dooren, H. M. Moody, R. Merkx, J. A. W. Kruijtzer, D. T. S. Rijkers, R. M. J. Liskamp, P. J. L. M. Quaedflieg, *Adv. Synth. Catal.* **2010**, *325*, 2399-2404.

[6] T. Nuijens, E. Piva, J. A. W. Kruijtzer, D. T. S. Rijkers, R. M. J. Liskamp, P. J. L. M. Quaedflieg, *Adv. Synth. Catal.* **2011**, *353*, 1039-1044.

[7] T. Nuijens, C. Cusan, A. C. H. M. Schepers, J. A. W. Kruijtzer, D. T. S. Rijkers, R. M. J. Liskamp, P. J. L. M. Quaedflieg, *J. Mol. Catal. B: Enzym.* **2011**, *71*, 79–84.

[8] (a) T. Miyazawa, S. Nakajo, M. Nishikawa, K. Hamahara, K. Imagawa, E. Ensatsu, R. Yanagihara, T. Yamada, *J. Chem. Soc. Perkin Trans. 1* **2001**, 82–86; (b) T. Miyazawa, K. Tanaka, E. Ensatsu, R. Yanagihari, T. Yamada, *J. Chem. Soc. Perkin Trans. 1* **2001**, 87-93; (c) S. M. A. Salam, K. Kagawa, T. Matsubara, K. Kawashiro, *Enzyme Microb. Technol.* **2008**, *43*, 537–543.

[9] D. F. Tai, *Curr. Org. Chem.* **2003**, *7*, 515-554.

[10] R. J. A. C. de Beer, B. Zarzycka, M. Mariman, H. I. V. Amatdjais-Groenen, M. J. Mulders, P. J. L. M. Quaedflieg, F. L. van Delft, S. B. Nabuurs, F. P. J. T. Rutjes, *ChemBioChem* **2012**, *13*, 1319-1326 .

[11] T. Miyazawa, E. Ensatsu, M. Hiramatsu, R. Yanagihara, T. Yamada, *J. Chem. Soc. Perkin Trans. 1* **2002**, 396-401.

[12] L. Abrahmsén, J. Tom, J. Burnier, K. A. Butcher, A. Kossiakoff, J. A. Wells, *Biochemistry* **1991**, *30*, 4151-4159.

[13] R. J. A. C. de Beer, T. Nuijens, L. Wiermans, P. J. L. M. Quaedflieg, F. P. J. T. Rutjes, *Org. Biomol. Chem.* **2012**. Accepted doi 10.1039/C2OB25662B

Chapter 7

Synthesis

The synthesis of all previously encountered mimetics and activating esters is described in this chapter. The reaction sequences primarily consisted of synthesis of the required alcohol, DCC-mediated esterification of this alcohol with an amino acid and subsequent deprotection.

7.1 Introduction

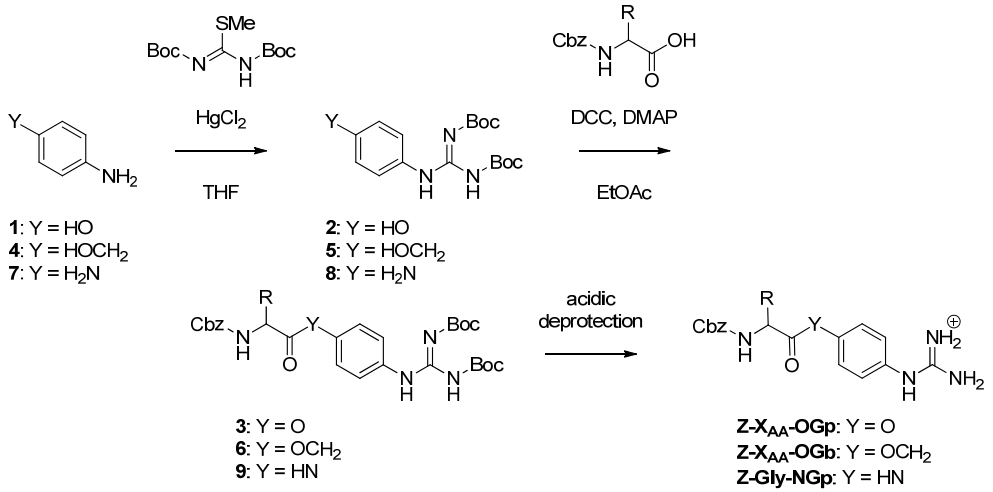
A variety of mimetics and activating esters were introduced and discussed in the previous chapters. The synthesis of these compounds was deliberately not described in the corresponding chapters as not to divert the attention from the experiments carried out with them. However, now the synthesis is considered, which can be regarded as a matrix that had to be filled, with the various alcohols on one axis and the various amino acids on the other. Initially, only the Z-Gly-OH esters were prepared from the various alcohols and, in a later stage, a range of amino acids were coupled to a selected set of alcohols. The syntheses are grouped based on similarity of the esters. To prevent overlap in the discussion, the typical issues encountered when working with amino acids with a functional side chain are described in § 7.2.2.

7.2 Aromatic guanidino- and amidino-based mimetics

7.2.1 Z-Gly-OGp, Z-Gly-OGb and Z-Gly-NGp

The synthesis of the aromatic guanidino group containing mimetics started from the corresponding amino alcohol (**1**,**4**), or aminoaniline (**7**) in the case of NGp (Table 7.1). First the diBoc protected guanidino group was installed (**2**, **5**, **8**), then a DCC-coupling with Z-Gly-OH was carried out to form the ester (**3**,**6**), or amide (**9**) in the case of NGp, followed by deprotection of the acid labile protecting groups.

Table 7.1 *Synthesis of aromatic guanidino-based mimetics*



Entry	Product	Yield (%)	Product	Yield (%)	Product	Yield (%)
1	**2**	65	**3G**	82	**Z-Gly-OGp**	quant
2	**5**	65	**6G**	97	**Z-Gly-OGb**	quant
3	**8**	55	**9G**	quant	**Z-Gly-NGp**	quant

In some cases spontaneous hydrolysis of the ester occurred under acidic conditions, resulting in a mixture that could not be separated because of the high polarity of the guanidine-containing compounds. Since hydrolysis was always less than 10 percent, and the uncoupled amino acid would not interfere in the enzymatic reactions, the crude product was used as such.

7.2.2 Z-X_{AA}-OGp and Z-X_{AA}-OGb

Both alcohols HOGp and HOGb were selected for coupling to a range of amino acids (Tables 7.2 and 7.3, respectively). The same scheme as depicted in Table 7.1 was followed for the synthesis, provided that functional amino acid side chains were properly protected. Acid labile protecting groups were preferred, as they could be simultaneously removed with the Boc groups, and are orthogonal to Cbz. Another aspect that was taken into account was the difficulty of purification after deprotection so that traceless protecting groups such as ^{t}Bu, Boc and TBS were chosen.

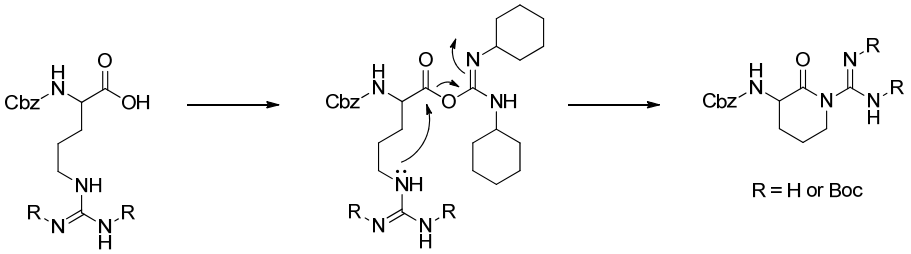
Table 7.2 *Synthesis of various Z-X_{AA}-OGp esters*

Entry	Product	Yield (%)	Product	Yield (%)
1	**3A**	63	**Z-L-Ala-OGp**	quant
2	**3a**	80	**Z-D-Ala-OGp**	quant
3	**3βA**	88	**Z-β-Ala-OGp**	quant
4	**3R(diBoc)**	85	**Z-L-Arg-OGp**	quant
5	**3N(Trt)**	81	**Z-L-Asn-OGp**	quant
6	**3E(O^{t}Bu)**	91	**Z-L-Glu-OGp**	quant
7	**3I**	97	**Z-L-Ile-OGp**	quant
8	**3F**	89	**Z-L-Phe-OGp**	quant
9	**3P**	82	**Z-L-Pro-OGp**	quant
10	**3T(OTBS)**	98	**Z-L-Thr-OGp**	quant

Table 7.3 *Synthesis of various Z-X_{AA}-OGb esters*

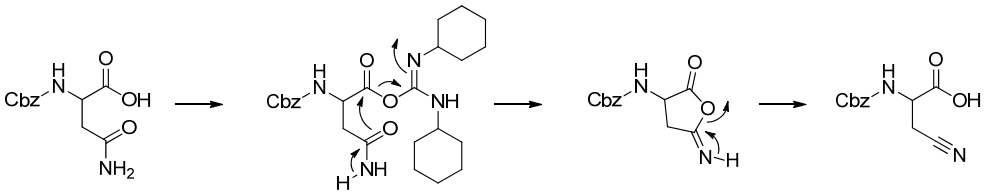
Entry	Product	Yield (%)	Product	Yield (%)
1	**6A**	99	**Z-L-Ala-OGb**	quant
2	**6a**	92	**Z-D-Ala-OGb**	quant
3	**6βA**	93	**Z-β-Ala-OGb**	quant
4	**6D(O^{t}Bu)**	89	**Z-L-Asp-OGb**	quant
5	**6Q(Trt)**	99	**Z-L-Gln-OGb**	82
6	**6I**	94	**Z-L-Ile-OGb**	quant
7	**6K(Boc)**	90	**Z-L-Lys-OGb**	65
8	**6F**	54	**Z-L-Phe-OGb**	quant
9	**6f**	88	**Z-D-Phe-OGb**	quant
10	**6P**	83	**Z-L-Pro-OGb**	quant

Our initial presumption was that Z-Arg-OH, when positively charged, could be used in unprotected form (entry 4 in Table 7.2, 7.7 and 7.10). However, attempts to form the ester under DCC coupling conditions mainly resulted in the cyclised δ-lactam form of Z-Arg-OH (Scheme 7.1).[1] Even diBoc protection could not completely prevent this side reaction, since approximately 15% of δ-lactam was observed.[2]



Scheme 7.1 *δ-lactam formation from Arg*

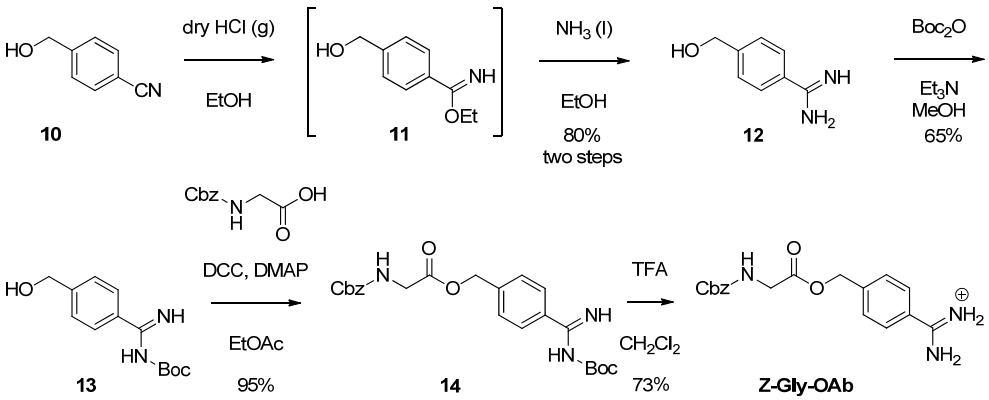
We neither anticipated that the amide-containing side chain of Z-Asn-OH would interfere with ester formation (entry 5 in Table 7.2, 7.7 and 7.10). Nevertheless, the major product appeared to be the dehydrated nitrile (Scheme 7.2).[3] After trityl protection, the desired product was obtained in good yield. The remainder of this protecting group was readily removed by washing the aqueous layer with Et_2O and subsequent lyophilisation.



Scheme 7.2 *Dehydration of Z-Asn-OH*

7.2.3 Z-Gly-OAb

The desired ester Z-Gly-OAb was accessed by the transformation of *p*-(hydroxymethyl)benzonitrile (**10**) into its so-called Pinner salt by subjection to a dry solution of HCl in ethanol (Scheme 7.3). Exclusion of water was essential to prevent unwanted hydrolysis of ethyl imidate **11** to the corresponding ester. Therefore, imidate **11** was obtained *in situ* and directly converted into amidine **12** by nucleophilic displacement of the ethoxide with ammonia dissolved in ethanol. Amidine **12** could not be coupled to Z-Gly-OH directly without side reactions, so that first the amidino group was mono Boc protected. After DCC coupling, the Boc group was removed again with TFA to obtain **Z-Gly-OAb** in an overall yield of 36% over five steps.



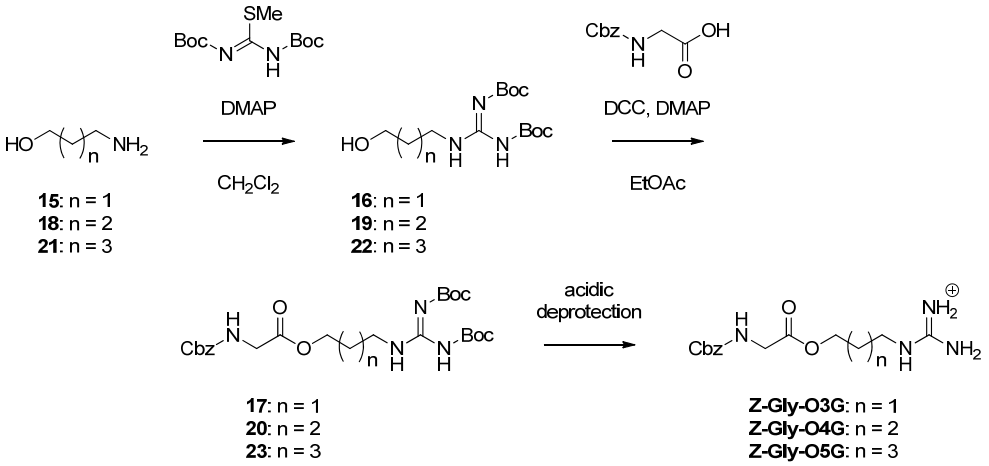
Scheme 7.3 *Synthesis of Z-Gly-OAb*

7.3 Aliphatic amino- and guanidino-based esters

7.3.1 Z-Gly-O3G, Z-Gly-O4G and Z-Gly-O5G

The synthesis of the aliphatic guanidino group-containing esters proceeded analogous to their aromatic counterparts (Table 7.4). Amino alcohols **15**, **18** and **21** were converted into the corresponding diBoc protected guanidino alcohols **16**, **19** and **22**, respectively, in good yields. Subsequent coupling with Z-Gly-OH and deprotection with TFA led to the targeted esters.

Table 7.4 *Synthesis of aliphatic guanidino-based esters*

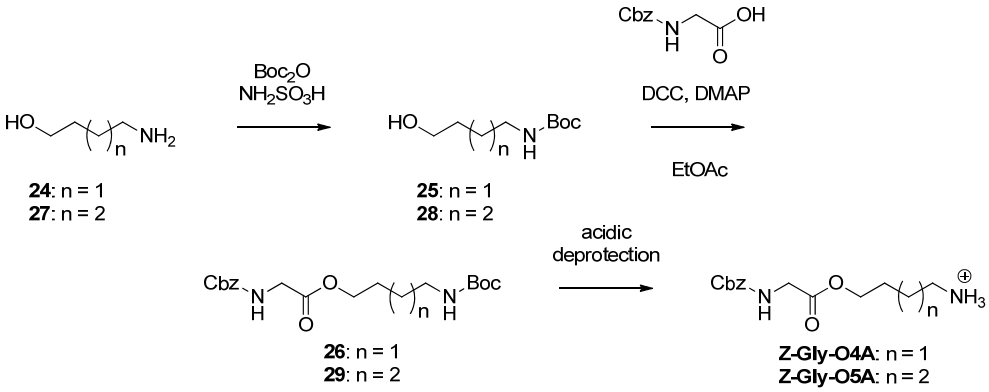


Entry	Product	Yield (%)	Product	Yield (%)	Product	Yield (%)
1	**16**	95	**17**	70	**Z-Gly-O3G**	quant
2	**19**	78	**20**	96	**Z-Gly-O4G**	quant
3	**22**	99	**23**	quant	**Z-Gly-O5G**	96

7.3.2 Z-Gly-O4A and Z-Gly-O5A

Aminobutanol and aminopentanol (**24** and **27**) were Boc-protected prior to coupling with Cbz-Gly-OH under the influence of DCC. Deprotection was performed with TFA in CH_2Cl_2 providing the amino esters **Z-Gly-O4A** and **Z-Gly-O5A** in excellent yields (Table 7.5).

Table 7.5 *Synthesis of aliphatic amino-based esters*



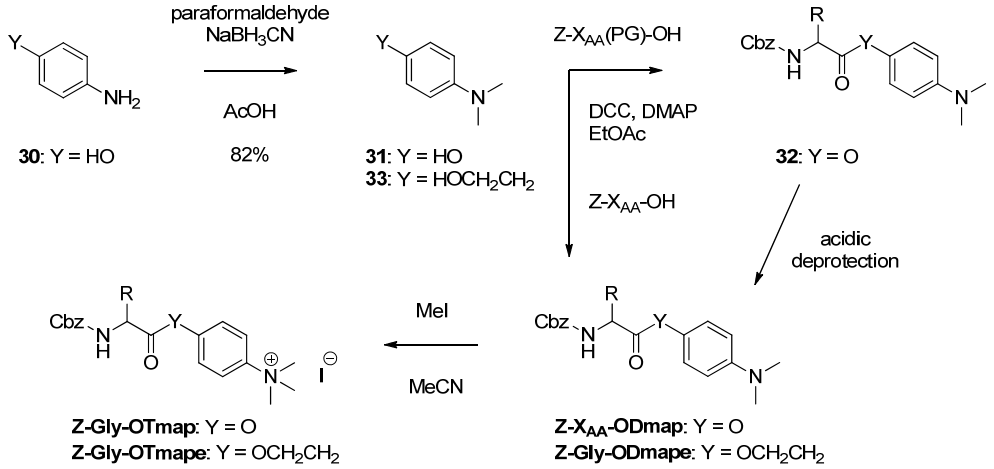
Entry	Product	Yield (%)	Product	Yield (%)	Product	Yield (%)
1	**25**	quant	**26**	83	**Z-Gly-O4A**	97
2	**28**	99	**27**	64	**Z-Gly-O5A**	quant

7.4 Dimethylamino- and trimethyl ammonium-containing esters

7.4.1 Z-Gly-ODmap, Z-Gly-OTmap, Z-Gly-ODmape and Z-Gly-OTmape

Dimethylamino phenol **31** was obtained via an Eschweiler-Clarke-like reaction of aminophenol **30** with paraformaldehyde and sodium cyanoborohydride in acetic acid (Table 7.6). The homologous phenethyl alcohol **33** was commercially available. DCC-coupling with Z-Gly-OH resulted in the desired esters **Z-Gly-ODmap** and **Z-Gly-ODmape**, which could then be converted into the corresponding trimethylammonium-containing esters with methyl iodide in MeCN providing **Z-Gly-OTmap** and **Z-Gly-OTmape**, respectively. In order to obtain a satisfactory yield, it is essential not to methylate prior to the DCC- coupling.

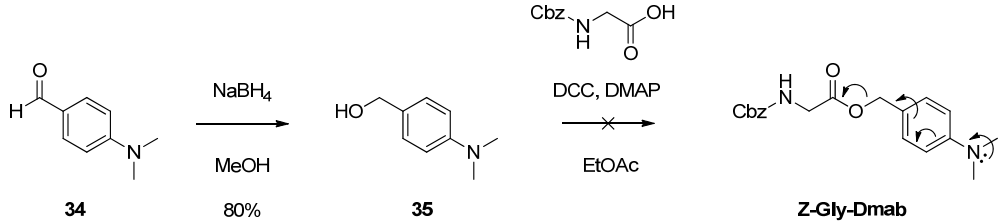
Table 7.6 *Synthesis of dimethylamino- and trimethyl ammonium-containing mimetics*



Entry	Product	Yield (%)	Product	Yield (%)
1	**Z-Gly-ODmap**	15	**Z-Gly-OTmap**	59
2	**Z-Gly-ODmape**	75	**Z-Gly-OTmape**	quant

7.4.2 Towards Z-Gly-ODmab

An attempt was also made to synthesise **Z-Gly-ODmab** (Scheme 7.4). 4-(Dimethylamino)benzyl alcohol **35** was prepared by the reduction of the corresponding aldehyde **34**. Subsequent DCC-mediated esterification with Z-Gly-OH seemed to work, but the purification was troublesome. Presumably, the formed product is an unstable ester that decomposed during work-up.



Scheme 7.4 *Synthesis towards Z-Gly-ODmab*

7.4.3 Z-X_{AA}-ODmap

Alcohol HODmap was selected to be coupled to a range of amino acids too (Table 7.7), which was done in the same way as depicted in the scheme that is part of Table 7.6. For amino acids with a functional side chain an additional deprotection step was required.

Table 7.7 *Synthesis of various Z-X_{AA}-ODmap esters*

Entry	Product	Yield (%)	Product	Yield (%)
1	**Z-L-Ala-ODmap**	29[a]	**n.a.**	n.a.
2	**Z-D-Ala-ODmap**	81[b]	**n.a.**	n.a.
3	**Z-β-Ala-ODmap**	92	**n.a.**	n.a.
4	**32R(diBoc)**	86	**Z-L-Arg-ODmap**	quant
5	**32N(Trt)**	40[b]	**Z-L-Asn-ODmap**	50
6	**32E(O^{t}Bu)**	97	**Z-L-Glu-ODmap**	quant
7	**Z-L-Ile-ODmap**	98	**n.a.**	n.a.
8	**Z-L-Phe-ODmap**	90	**n.a.**	n.a.
9	**Z-L-Pro-ODmap**	88	**n.a.**	n.a.
10	**32T(O^{t}Bu)**	86	**Z-L-Thr-ODmap**	quant

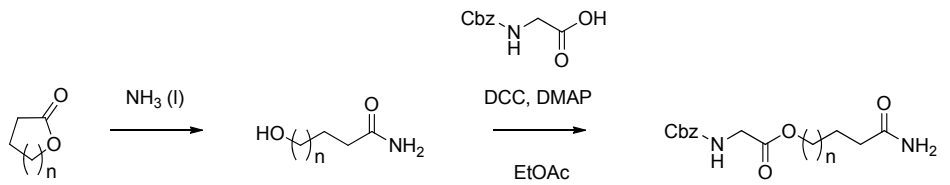
[a] Pure fraction from first column, no second column purification of remainder
[b] T3P[4] was used as coupling reagent

7.5 Miscellaneous esters

7.5.1 Z-Gly-O3Cam and Z-Gly-O4Cam

These two OCam esters were synthesised by ring-opening of γ-butyrolactone (**36**) and δ-valerolactone (**38**) with liquid ammonia in a sealed tube at 40 °C and the resulting hydroxyl amides (**37** and **39**) were coupled with Z-Gly-OH using EDC·HCl or DCC.

Table 7.8 *Synthesis of carbamoyl-based esters*



36: n = 1
38: n = 2

37: n = 1
39: n = 2

Z-Gly-O3Cam: n = 1
Z-Gly-O4Cam: n = 2

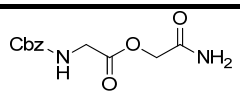
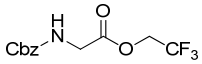
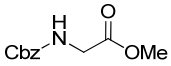
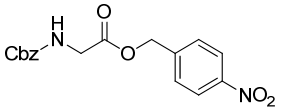
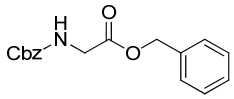
Entry	Product	Yield (%)	Product	Yield (%)
1	**35**	97	**Z-Gly-O3Cam**	46
2	**37**	62	**Z-Gly-O4Cam**	94

7.5.2 Z-Gly-OCam, Z-Gly-OTfe, Z-Gly-OMe, Z-Gly-ONb and Z-Gly-OBn

The esters in this section were all obtained in a single reaction step (Table 7.9). To prepare **Z-Gly-OCam**, Z-Gly-OH was converted into its potassium salt and reacted with chloroacetamide in DMF.[5] Z-Gly-OH was coupled to trifluoroethanol with EDC to afford **Z-Gly-OTfe**. **Z-Gly-OMe** was obtained after a reaction of Z-Gly-OH with methyl chloroformate. Z-Gly-OH was coupled to *p*-nitrobenzyl alcohol with DDC to obtain **Z-Gly-**

ONb. Benzyl bromide in DMF under basic conditions was used to prepare **Z-Gly-OBn** from Z-Gly-OH.

Table 7.9 *Synthesis of miscellaneous esters*

Entry	Structure	Product	Yield (%)
1		**Z-Gly-OCam**	49
2		**Z-Gly-OTfe**	64
3		**Z-Gly-OMe**	96
4		**Z-Gly-ONb**	83
5		**Z-Gly-OBn**	84

7.5.3 Z-X_{AA}-OBn

Besides Z-Gly-OH, a range of other amino acids was treated with benzyl bromide to obtain the corresponding **Z-X_{AA}-OBn** esters (Table 7.10). In case amino acids with a protected functional side chain were applied (**40**), acidic deprotection was necessary to afford the final products.

Table 7.10 *Synthesis of various Z-X_{AA}-OBn esters*

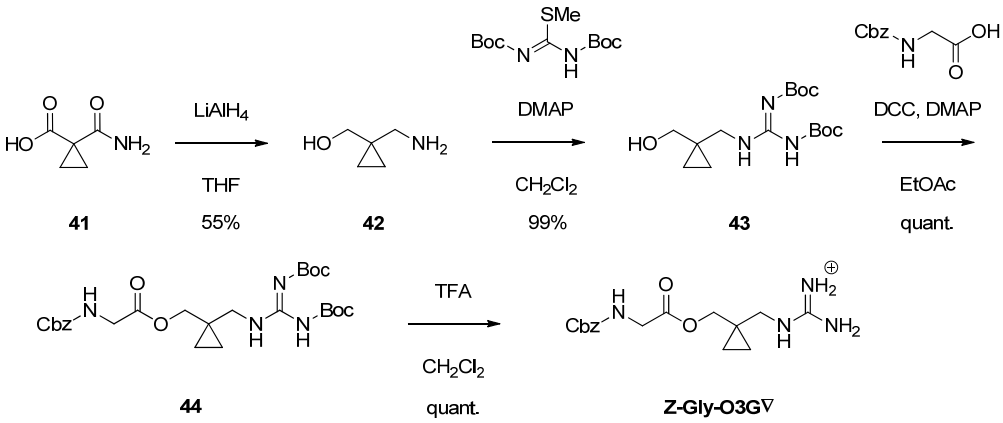
Entry	Product	Yield (%)	Product	Yield (%)
1	**Z-L-Ala-OBn**	89	**n.a.**	n.a.
2	**Z-D-Ala-OBn**	97	**n.a.**	n.a.
3	**Z-β-Ala-OBn**	93	**n.a.**	n.a.
4	**40R(diBoc)**	92	**Z-L-Arg-OBn**	95
5	**40N(Trt)**	82	**Z-L-Asn-OBn**	61
6	**40E(O^{t}Bu)**	85	**Z-L-Glu-OBn**	78
7	**Z-L-Ile-OBn**	91	**n.a.**	n.a.
8	**Z-L-Phe-OBn**	n.a.[a]	**n.a.**	n.a.
9	**Z-L-Pro-OBn**	98	**n.a.**	n.a.
10	**Z-L-Thr-OBn**	n.a.[a]	**n.a.**	n.a.

[a] Commercially available

7.6 O3G variants

7.6.1 Z-Gly-O3G∇

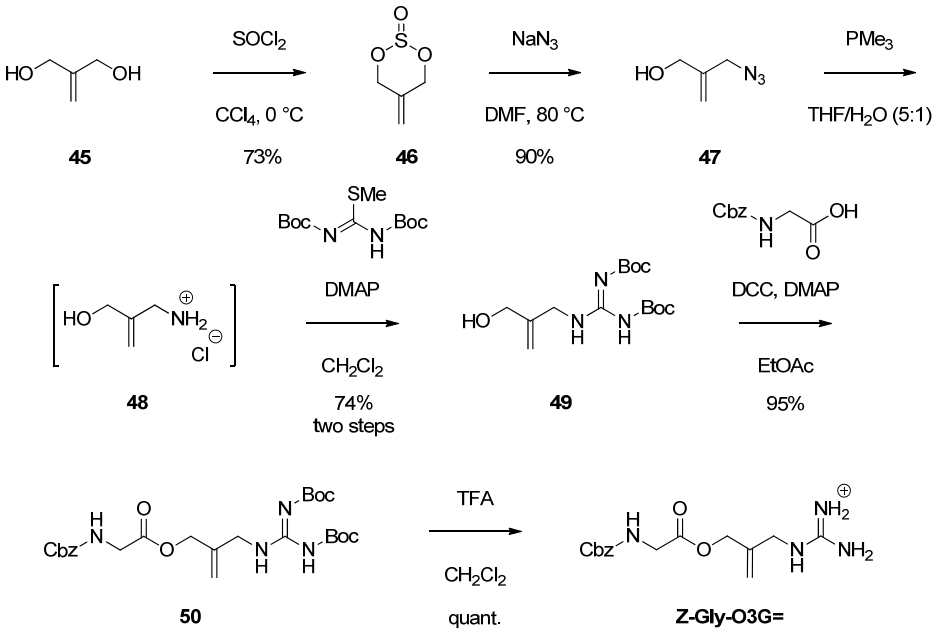
The synthesis of **Z-Gly-O3G∇** commenced with $LiAlH_4$ reduction of acid amide **41** to provide amino alcohol **42** in 55% yield (Scheme 7.5).[6] This low value is mainly due to the difficult work-up. Subsequent guanylation of the amine (**43**) and esterification of the alcohol resulted in Boc-protected precursor **44** in excellent yield. Acidic deprotection afforded **Z-Gly-O3G∇** quantitatively as the TFA salt.



Scheme 7.5 *Synthesis of Z-Gly-O3G∆*

7.6.2 Z-Gly-O3G=

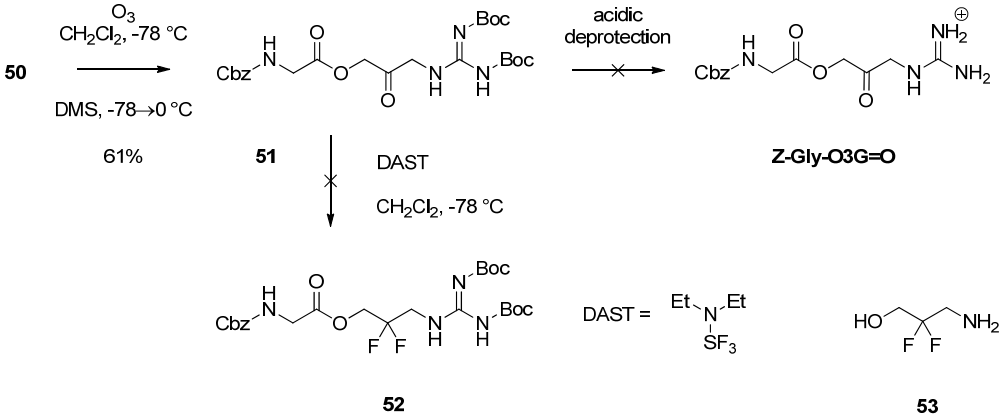
Diol **45**, with the desired methylene group already present, was taken as the starting compound for the synthesis of **Z-Gly-O3G=** (Scheme 7.6). In order to convert only one of the alcohols into an amine in a more than statistical distribution, it was first turned into cyclic sulfite **46**, followed by ring-opening with sodium azide to acquire azido alcohol **47**.[7] Staudinger reduction with trimethylphosphine followed by acidic work-up yielded amino alcohol **48**, which was directly guanylated (**49**). Subsequent coupling with Z-Gly-OH (**50**) and deprotection with TFA gave rise to **Z-Gly-O3G=** in 46% overall yield.



Scheme 7.6 *Synthesis of Z-Gly-O3G=*

7.6.3 Towards Z-Gly-O3G=O and Z-Gly-O3GF_2

Starting from **50**, the corresponding ketone was obtained *via* ozonolysis (Scheme 7.7). Besides the major product **51**, many byproducts were formed. In addition, the subsequent deprotection step appeared troublesome; treatment with both TFA in CH_2Cl_2 or HCl in dioxane led to complete decomposition of ketone **51**. Hence, it was concluded that **Z-Gly-O3G=O** is an unstable compound.



Scheme 7.7 *Synthesis towards Z-Gly-O3G=O and Z-Gly-O3GF_2*

The synthesis of **Z-Gly-O3GF$_2$** was attempted by utilising DAST as nucleophilic fluoride donor on ketone **50** (Scheme 7.8). Unfortunately, no reaction occurred. Alternative approaches, involving the synthesis of fluorinated amino alcohol **53** by various methods, failed to succeed.

7.7 Nucleophiles

Some of the nucleophiles used in the enzymatic experiments, were not commercially available and thus had to be synthesised. A convenient method for the synthesis of amino acid *p*-nitroanilides has been described by Rijkers *et al.*[8] Table 7.11 shows the results of this coupling strategy using $POCl_3$ and acidic deprotection.

Table 7.11 *Synthesis of pNA-containing nucleophiles*

Entry	Product	Yield (%)	Product	Yield (%)
1	**Boc-Pro-pNA**	65	**Pro-pNA**	94
2	**Boc-Ser(O^{t}Bu)-pNA**	46	**Ser-pNA**	85

7.8 Conclusions

The synthesis of most mimetics and activating esters appeared relatively straightforward. Exceptions were the Z-Gly-O3G variants, which however, were mainly synthesised to fundamentally analyse a principle and not with a large-scale application in mind. In general, no attempts were made to improve the yields, since small amounts of product were sufficient to carry out the enzymatic experiments.

7.9 Acknowledgements

Helene Amatdjais-Groenen (OGb compounds in § 7.2.2, compounds in § 7.3 except for 16, 17 and Z-Gly-O3G, Z-Gly-ONb in § 7.5.2, 40R, N and E and corresponding deprotected OBn esters in § 7.5.3), Michiel Mariman (compounds in § 7.4.1 and 7.4.2 and Z-Gly-OCam, Z-Gly-OTfe and Z-Gly-OMe in § 7.5.2), Sander Jans (preliminary investigation to OGp compounds in § 7.2.2), Berry Bögels (Z-Gly-O3Cam and precursor 37 in § 7.5.1, compounds in § 7.2.3 and § 7.6) and Marc Mulders (all compounds in § 7.4.3 except for 32R and Z-Arg-ODmap which were prepared by HA) are gratefully acknowledged for their contribution to the synthesis.

7.10 Experimental section

General remarks:

All chemicals were obtained from commercial sources and used without further purification, unless stated otherwise. If appropriate, reactions were carried out under an inert atmosphere of dry nitrogen or argon. Standard syringe techniques were applied for the transfer of dry solvents and air- or moisture-sensitive reagents. Reactions were followed and R_f values were obtained using thin layer chromatography (TLC) on silica gel-coated plates (Merck 60 F254) with the

indicated solvent mixture. Detection was performed with UV-light, and/or by charring at ~150 °C after dipping into aqueous basic permanganate. Melting points were analysed with a Büchi melting point B-545. IR spectra were recorded on an ATI Mattson Genesis Series FTIR spectrometer, or a Bruker Tensor 27 FTIR spectrometer. NMR spectra were recorded on a Bruker DMX 300 (300 MHz and 75 MHz for ^{1}H and ^{13}C, respectively). ^{1}H-NMR chemical shifts are reported in parts per million (ppm) relative to tetramethylsilane (TMS) as internal standard, or a residual proton peak of the solvent: δ = 7.26 ppm for $CDCl_3$, δ = 3.31 ppm for CD_3OD, δ = 2.94 ppm for CD_3CN and δ = 2.50 ppm for DMSO-d_6. Multiplicities are reported as: s (singlet), d (doublet), t (triplet), q (quartet), dd (doublet of doublets), dt (doublet of triplets), dq (double quartet), ddd (double, double doublet), ddt (double, double triplet) or m (multiplet). Broad peaks are indicated by br. Coupling constants are reported as *J*-values in Hz. The number of protons (n) for a given resonance is indicated as nH, and is based on spectral integration values. ^{13}C-NMR chemical shifts (δ) are reported in ppm relative to $CDCl_3$ (δ = 77.0), CD_3OD (δ = 49.0), CD_3CN (δ = 1.24) or DMSO-d_6 (δ = 39.5). Column or flash chromatography was carried out using ACROS silica gel (0.035-0.070 mm, and ca 6 nm pore diameter). Optical rotations were determined with a Perkin Elmer 241 polarimeter. High resolution mass spectra were recorded on a JEOL AccuTOF (ESI), or a MAT900 (EI, CI, and ESI). Elemental analyses were carried out using a Carlo Erba Instruments CHNS-O EA 1108 element analyser.

General procedure A1: DCC-coupling of Cbz-protected amino acids with an alcohol

To a cooled solution (0 °C) of Z-X_{AA}-OH (1.4 equiv), alcohol (1 equiv) and DMAP (0.2 equiv) in EtOAc (10 mL), DCC (1.4 equiv) was added slowly. The reaction mixture was stirred at 0 °C for 1 h and an additional 2 h at room temperature. The solid DCU was filtered off and in some experiments a washing step was performed (**general procedure A2**) before the solvent was evaporated *in vacuo.* In that case the reaction mixture was poured in citric acid (5%, 5 mL) and extracted with EtOAc (3 × 25 mL). The combined organic layers were washed with saturated $NaHCO_3$ (5 mL) and brine (10 mL), dried over sodium sulfate and then evaporated to dryness. The crude reaction mixture was purified with column chromatography.

General procedure A3: EDC coupling of Cbz-protected amino acid with an alcohol or amino amide

EDC (1.2 equiv) was added to a cooled solution (0 °C) of alcohol or amino amide (1 equiv), Z-X_{AA}-OH (1 equiv) and DMAP or DIPEA (1 equiv) in CH_2Cl_2. The solution was stirred at room temperature for at least 3 hours. The mixture was concentrated and the residue was dissolved in EtOAc, and washed with water (3 × 10 mL). The organic layer was dried over Na_2SO_4, concentrated and purified by column chromatography if the ^{1}H NMR showed impurities.

General procedure B1: Boc-deprotection with TFA

The Boc-protected compound (100 mg) was dissolved in CH_2Cl_2 (2 mL) to which TFA was added (0.5 mL). The reaction mixture was stirred at room temperature overnight. The solvents were removed under reduced pressure and co-evaporated with ^{t}BuOH (3 × 10 mL). The obtained oil was lyophilized from a H_2O/dioxane mixture (10 mL) in the presence of 2M HCl (0.5 mL), to give the product. In several experiments co-evaporation took place with CH_2Cl_2 and the remaining oily TFA salt was directly used (**general procedure B2**).

General procedure B3: Boc-deprotection with HCl in dioxane

The Boc-protected compound (100 mg) was treated with a commercially available solution of ~ 4 M HCl in dioxane (2 mL) and stirred at room temperature overnight. The solvents were removed under reduced pressure and co-evaporated with ^{t}BuOH (3 × 10 mL). The obtained oil was lyophilized from a H_2O/dioxane mixture (10 mL) to give the product.

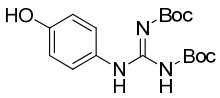
General procedure C: benzyl ester synthesis of Cbz-protected amino acids

To a solution of Z-X_{AA}-OH (1 equiv) in DMF (1.5 mL) was added K_2CO_3 (1 equiv) and benzyl bromide (1.1 equiv). After being stirred for 6 hours the reaction mixture was quenched with water (10 mL) and extracted with EtOAc/heptane 1:1 (3 × 25 mL). The combined organic layers were washed with water (3 × 50 mL) and brine (1 × 50 mL), dried over Na_2SO_4 and then evaporated to dryness. The crude reaction mixture was purified with column chromatography.

General procedure D: *p*-nitroanilide synthesis with $POCl_3$[8]

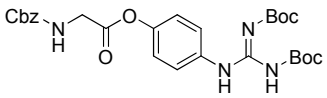
A protected amino acid (1 equiv) and recrystallized *p*-nitroaniline (1 equiv) were dissolved in dry pyridine. The clear yellowish solution was cooled to -15 °C and phosphorus oxychloride (1.1 equiv) was added dropwise with vigorous stirring. During the addition, the reaction mixture coloured deeply red and became turbid in the course of 5 to 20 minutes. The colour of the suspension slowly changed to orange, the reaction being complete after a total of 10 to 30 min (monitored by TLC). The reaction mixture was then quenched with crushed ice/water (100 mL) and the nitroanilide was extracted into EtOAc (1 × 50 mL and 3 × 30 mL). The combined EtOAc layers were washed with saturated $NaHCO_3$ and brine (3 × 30 mL each). After drying on Na_2SO_4, the EtOAc layer was filtered and evaporated *in vacuo*. The residue was co-evaporated successively with toluene, EtOAc and MeOH.

_p_-[_N'_, _N''_-di(Boc)guanidino]phenol (2)



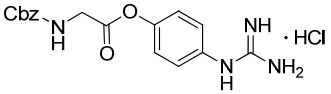
N,N'-di(Boc)-*S*-methylisothiourea (2.90 g, 10.0 mmol, 1 equiv) and *p*-aminophenol (1.64 g, 15.0 mmol, 1.5 equiv) were dissolved in dry THF (60 mL) and this mixture was cooled to 0 °C before $HgCl_2$ (2.99 g, 11.0 mmol, 1.1 equiv) was added. After stirring for 20 min under argon, the temperature was raised to room temperature and the mixture was stirred for 20 h. The white precipitate that was formed during the reaction was filtered off and the filtrate was concentrated under reduced pressure. Recrystallisation from methanol yielded 1.53 g (43%) of the pure product. The mother liquor was then evaporated to dryness and the remaining solid was purified by column chromatography (EtOAc/heptane 1:9→1:2) to afford **3** (753 mg, 65%) as an off-white solid. R_f 0.37 (EtOAc/heptane 1:2). Mp dec at >240 °C. IR (film) 3264, 2979, 2737, 1720, 1647, 1517, 1409, 1227, 1152, 1112, 1059 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.61 (s, NH), 10.04 (s, NH), 7.14 (d, *J* = 8.6 Hz, 2H), 6.68 (d, *J* = 8.3 Hz, 2H), 6.22 (br s, OH), 1.54 (s, 9H), 1.46 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 163.2, 155.9, 155.3, 153.2, 127.0, 126.3, 116.1, 83.7, 79.8, 28.1. HRMS (ESI) *m/z* calcd for $C_{17}H_{26}N_3O_5$ $(M+H)^+$: 352.1873, found: 352.1875. Spectral data were in accordance with those reported in literature.[9]

N^{α}-Cbz-Glycine _p_-[_N'_,_N''_-di(Boc)guanidino]phenyl ester (3G)



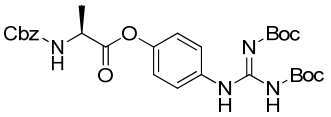
General procedure A1 using Z-Gly-OH (879 mg, 4.2 mmol), **2** (1.05 g, 3.00 mmol), DMAP (73 mg, 0.6 mmol) and DCC (867 mg, 4.21 mmol). The product was obtained as a white solid after purification by column chromatography (MeOH in CH_2Cl_2, 1→4%) (1.33 g, 82%). R_f 0.67 (4% MeOH in CH_2Cl_2). Mp 111 °C. IR (film) 2978, 2928, 1779, 1720, 1640, 1508, 1412, 1240, 1154, 1114, 1057 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.62 (s, NH), 10.35 (s, NH), 7.62 (d, *J* = 8.5 Hz, 2H), 7.41-7.28 (m, 5H), 7.07 (d, *J* = 8.3 Hz, 2H), 5.34 (m, NH), 5.15 (s, 2H), 4.22 (d, *J* = 5.3 Hz, 2H), 1.53 (s, 9H), 1.50 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 168.6, 163.4, 156.2, 153.5, 153.3, 146.8, 136.1, 134.8, 128.6, 128.2, 128.1, 123.2, 121.6, 83.9, 79.8, 67.3, 42.9, 28.2, 28.1. HRMS (ESI) *m/z* calcd for $C_{27}H_{34}N_4NaO_8$ $(M+Na)^+$: 565.2274, found: 565.2274.

N^{α}-Cbz-Glycine *p*-guanidinophenyl ester (Z-Gly-OGp)



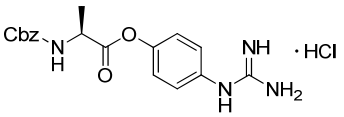
General procedure B1 using **3G** (100 mg, 0.184 mmol). The product was obtained as a sticky oil (69.8 mg, quant). R_f 0.50 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). IR (film) 3309, 3166, 2950, 1770, 1706, 1671, 1629, 1588, 1507, 1455, 1280, 1201, 1166, 1053 cm^{-1}. ^1H NMR (DMSO-d_6, 300 MHz): δ 10.03 (s, NH), 7.89 (t, *J* = 5.9 Hz, NH), 7.56 (br s, 4NH), 7.40-7.25 (m, 7H), 7.21-7.15 (m, 2H), 5.08 (s, 2H), 4.07 (d, *J* = 6.1 Hz, 2H). ^{13}C NMR (DMSO-d_6, 75 MHz): δ 169.0, 156.5, 156.0, 148.2, 136.8, 132.9, 128.3, 127.8, 127.7, 125.7, 122.7, 65.6, 42.4. HRMS (ESI) *m/z* calcd for $C_{17}H_{18}N_4NaO_4$ $(M+Na)^+$: 365.1226, found: 365.1231.

N^{α}-Cbz-L-Alanine *p*-[*N',N''*-di(Boc)guanidino]phenyl ester (3A)



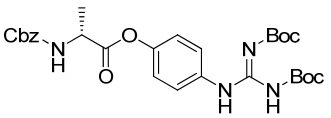
General procedure A1 using Z-L-Ala-OH (625 mg, 2.8 mmol), **2** (703 mg, 2.00 mmol), DMAP (49 mg, 0.41 mmol) and DCC (578 mg, 2.81 mmol). The product was obtained as a white solid after purification by column chromatography (MeOH in CH_2Cl_2, 1→2%) (702 mg, 63%). R_f 0.63 (2% MeOH in CH_2Cl_2). Mp 60 °C. $[\alpha]_D^{20}$ –14.7 (*c* 0.30, CH_2Cl_2). IR (film) 2976, 2933, 1761, 1717, 1634, 1507, 1410, 1238, 1150, 1113, 1058 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.63 (s, NH), 10.33 (s, NH), 7.60 (d, *J* = 8.5 Hz, 2H), 7.39-7.29 (m, 5H), 7.03 (d, *J* = 8.1 Hz, 2H), 5.53 (d, *J* = 7.0 Hz, NH), 5.12 (s, 2H), 4.67-4.52 (m, 1H), 1.57-1.46 (m, 21H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.5, 163.3, 155.6, 153.4, 153.1, 146.9, 136.1, 134.5, 128.4, 128.1, 128.0, 123.1, 121.4, 83.7, 79.6, 66.8, 49.7, 28.0, 27.9, 18.2. HRMS (ESI) *m/z* calcd for $C_{28}H_{37}N_4O_8$ $(M+H)^+$: 557.2611, found: 557.2605.

N^{α}-Cbz-L-Alanine *p*-guanidinophenyl ester (Z-L-Ala-OGp)



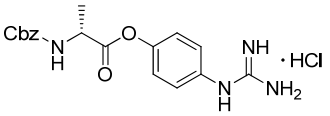
General procedure B1 using **3A** (100 mg). The product was obtained as a sticky oil (quant). R_f 0.50 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). $[\alpha]_D^{20}$ –34.5 (*c* 0.24, MeOH). IR (film) 3313, 3175, 2941, 1757, 1704, 1671, 1630, 1589, 1507, 1452, 1257, 1195, 1169, 1067 cm^{-1}. ^1H NMR (CD_3CN, 300 MHz): δ 9.87 (s, NH), 7.40-7.25 (m, 7H), 7.18-7.10 (m, 2H), 7.02 (br s, 3NH), 6.31 (d, *J* = 6.6 Hz, NH), 5.09 (s, 2H), 4.43 (dq, *J* = 7.3, 7.3 Hz, 1H), 1.50 (d, *J* = 7.3 Hz, 3H). ^{13}C NMR (CD_3CN, 75 MHz): δ 172.7, 158.1, 157.0, 150.7, 133.0, 129.4, 128.9, 128.7, 128.0, 124.0, 67.2, 51.0, 17.4. HRMS (ESI) *m/z* calcd for $C_{18}H_{20}N_4NaO_4$ $(M+Na)^+$: 379.1382, found: 379.1382.

N^{α}-Cbz-D-Alanine p-[*N',N''*-di(Boc)guanidino]phenyl ester (3a)



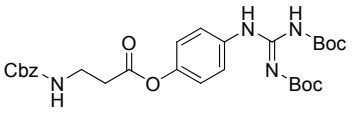
General procedure A1 using Z-D-Ala-OH (313 mg, 1.40 mmol), **2** (351 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (MeOH in CH_2Cl_2, 1→3%) (445 mg, 80%). R_f 0.67 (2% MeOH in CH_2Cl_2). Mp 60 °C. $[\alpha]_D^{20}$ +16.4 (*c* 0.32, CH_2Cl_2). IR (film) 2976, 2941, 1766, 1719, 1639, 1511, 1412, 1239, 1152, 1114, 1059 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.63 (s, NH), 10.34 (s, NH), 7.61 (d, *J* = 8.4 Hz, 2H), 7.44-7.27 (m, 5H), 7.04 (d, *J* = 8.1 Hz, 2H), 5.46 (d, *J* = 6.7 Hz, NH), 5.13 (s, 2H), 4.67-4.53 (m, 1H), 1.57-1.46 (m, 21H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.6, 163.4, 155.6, 153.4, 153.2, 146.9, 136.1, 134.6, 128.5, 128.1, 128.0, 123.1, 121.5, 83.7, 79.6, 66.9, 49.8, 28.1, 28.0, 18.4. HRMS (ESI) *m/z* calcd for $C_{28}H_{37}N_4O_8$ $(M+H)^+$: 557.2611, found: 557.2607.

N^{α}-Cbz-D-Alanine *p*-guanidinophenyl ester (Z-D-Ala-OGp)



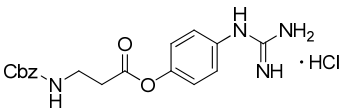
General procedure B1 using **3a** (100 mg). The product was obtained as a sticky oil (quant). R_f 0.50 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). $[\alpha]_D^{20}$ +33.3 (*c* 0.23, MeOH). IR (film) 3311, 3175, 2940, 1761, 1701, 1672, 1631, 1588, 1507, 1453, 1258, 1196, 1169, 1067 cm^{-1}. ^1H NMR (CD_3CN, 300 MHz): δ 9.89 (s, NH), 7.41-7.24 (m, 7H), 7.18-7.10 (m, 2H), 7.03 (br s, 4NH), 6.30 (d, *J* = 6.5 Hz, NH), 5.09 (s, 2H), 4.43 (dq, *J* = 7.2, 7.3 Hz, 1H), 1.50 (d, *J* = 7.3 Hz, 3H). ^{13}C NMR (CD_3CN, 75 MHz): δ 172.7, 165.9, 158.1, 157.0, 150.7, 133.0, 129.4, 129.1, 128.9, 128.8, 128.7, 128.0, 124.0, 67.2, 51.0, 17.4. HRMS (ESI) *m/z* calcd for $C_{18}H_{20}N_4NaO_4$ $(M+Na)^+$: 379.1382, found: 379.1395.

N^{α}-Cbz-β-Alanine *p*-[*N',N''*-di(Boc)guanidino]phenyl ester (3βA)



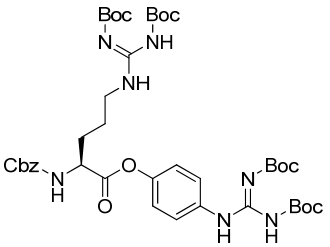
General procedure A1; using Z-β-Ala-OH (313 mg, 1.40 mmol), **2** (351 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (MeOH in CH_2Cl_2, 1→3%) (490 mg, 88%). R_f 0.60 (2% MeOH in CH_2Cl_2). Mp 56 °C. IR (film) 2980, 2933, 1766, 1717, 1630, 1502, 1409, 1237, 1149, 1113, 1057, 731cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.62 (s, NH), 10.33 (s, NH), 7.65-7.56 (m, 2H), 7.42-7.28 (m, 5H), 7.10-6.99 (m, 2H), 5.39-5.29 (m, NH), 5.11 (s, 2H), 3.55 (dt, *J* = 5.8, 5.9 Hz, 2H), 2.79 (t, *J* = 5.6 Hz, 2H), 1.53 (s, 9H), 1.50 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 170.8, 163.4, 156.2, 153.5, 153.2, 146.9, 136.4, 134.5, 128.5, 128.1, 128.0, 123.1, 121.7, 83.8, 79.7, 66.7, 36.4, 34.6, 28.1, 28.0. HRMS (ESI) *m/z* calcd for $C_{28}H_{37}N_4O_8$ $(M+H)^+$: 557.2611, found: 557.2604.

N^{α}-Cbz-β-Alanine *p*-guanidinophenyl ester (Z-β-Ala-OGp)



General procedure B1 using **3βA** (100 mg). The product was obtained as a sticky oil (quant). R_f 0.50 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). IR (film) 3330, 3175, 3062, 2946, 1754, 1701, 1672, 1629, 1588, 1508, 1455, 1257, 1200, 1166, 1065, 1014 cm^{-1}. ^1H NMR (CD_3CN, 300 MHz): δ 9.82 (s, NH), 7.40-7.25 (m, 7H), 7.20 (m, 2H), 6.98 (br s, 3NH), 5.94 (br s, NH), 5.07 (s, 2H), 3.47 (dt, *J* = 6.4, 6.4 Hz, 2H), 2.76 (t, *J* = 6.5 Hz, 2H). ^{13}C NMR (CD_3CN, 75 MHz): δ 171.4, 158.1, 157.3, 150.8, 132.8, 129.4, 128.8, 128.6, 128.0, 124.3, 66.9, 37.5, 35.5. HRMS (ESI) *m/z* calcd for $C_{19}H_{20}N_4O_4$ $(M+H)^+$: 357.1563, found: 357.1546.

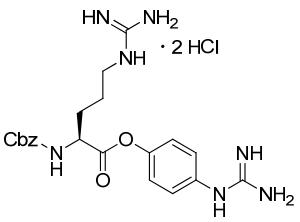
N^{α}-Cbz-L-Arginine(diBoc) *p*-[*N',N''*-di(Boc)guanidino]phenyl ester (3R)



General procedure A1 using Z-L-Arg(diBoc)-OH (712 mg, 1.40 mmol), **2** (351 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:9→1:2) (718 mg, 85%). R_f 0.70 (2% MeOH in CH_2Cl_2). Mp 84 °C. $[\alpha]_D^{20}$ +3.6 (*c* 0.98, CH_2Cl_2). IR (film) 2984, 2937, 1766, 1720, 1640, 1511, 1414, 1334, 1240, 1156, 1056 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.61 (s, NH), 11.48 (s, NH), 10.34 (s, NH), 8.39 (t, *J* = 5.4 Hz, NH), 7.62 (d, *J* = 8.9 Hz, 2H), 7.39-7.29 (m, 5H), 7.04 (d, *J* = 8.6 Hz, 2H), 5.70 (d, *J* = 8.0 Hz, NH), 5.14 (s, 2H), 4.68-4.56 (m, 1H), 3.61-3.37 (m, 2H), 2.14-1.95 (m, 1H), 1.83-1.67 (m, 3H), 1.54 (s, 9H), 1.50 (s, 9H), 1.49 (s, 9H), 1.48 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 170.8, 163.5, 156.3, 153.5, 153.3, 146.9, 134.8, 128.5, 128.1, 123.2, 121.6, 83.8, 83.2, 79.7, 79.4, 67.1, 54.0,

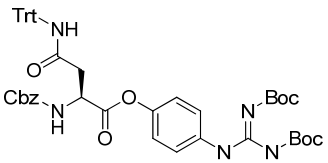
40.2, 29.4, 28.3, 28.2, 28.1, 25.5. HRMS (ESI) *m/z* calcd for $C_{41}H_{61}N_7O_{12}$ $(M+2H)^+$: 843.4378, found: 843.4373.

N^{α}-Cbz-L-Arginine *p*-guanidinophenyl ester (Z-L-Arg-OGp)



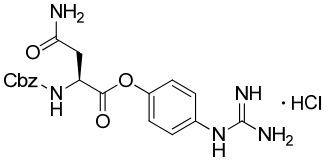
General procedure B1 using **3R** (100 mg). The product was obtained as a sticky oil (quant). R_f 0.50 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). $[\alpha]_D^{20}$ –22.1 (*c* 0.40, MeOH). IR (film) 3328, 3165, 2958, 1757, 1668, 1589, 1507, 1262, 1197, 1165 cm^{-1}. ^1H NMR (DMSO-d_6, 300 MHz): δ 10.15 (s, NH), 8.04 (d, *J* = 7.1 Hz, NH), 7.94 (t, *J* = 5.4 Hz, NH), 7.60 (br s, 4NH), 7.42-7.25 (m, 7H), 7.19-7.11 (m, 2H), 5.08 (s, 2H), 4.34-4.23 (m, 1H), 3.16 (dt, *J* = 6.3, 6.3 Hz, 2H), 2.02-1.56 (m, 6H). ^{13}C NMR (DMSO-d_6, 75 MHz): δ 170.9, 157.0, 156.2, 156.1, 148.2, 136.7, 132.9, 128.3, 127.8, 127.7, 125.6, 122.6, 65.6, 53.7, 31.2, 27.5, 25.0, 21.0. HRMS (ESI) *m/z* calcd for $C_{21}H_{27}N_7O_4$ $(M+H)^+$: 442.2203, found: 442.2188.

N^{α}-Cbz-L-Asparagine(Trt) *p*-[*N',N''*-di(Boc)guanidino]phenyl ester (3N)



General procedure A1 using Z-L-Asn(Trt)-OH (356 mg, 0.70 mmol), **2** (176 mg, 0.50 mmol), DMAP (12 mg, 0.10 mmol) and DCC (144 mg, 0.70 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:9→1:2) (339 mg, 81%). R_f 0.18 (EtOAc/heptane 1:5). Mp 94 °C. $[\alpha]_D^{20}$ +17.8 (*c* 0.85, CH_2Cl_2). IR (film) 2976, 2924, 1766, 1719, 1640, 1506, 1412, 1239, 1153, 1114, 1057, 699 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.61 (s, NH), 10.30 (s, NH), 7.54 (d, *J* = 9.0 Hz, 2H), 7.37-7.12 (m, 20H), 6.76 (d, *J* = 9.0 Hz, 2H), 6.65 (s, NH), 6.11 (d, *J* = 9.0 Hz, NH), 5.14 (s, 2H), 4.82-4.72 (m, 1H), 3.24 (dd, *J* = 3.7, 16.4, 1H), 2.96 (dd, *J* = 3.5, 16.1, 1H), 1.54 (s, 9H), 1.52 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 169.9, 169.1, 163.5, 156.2, 153.5, 153.3, 147.2, 144.2, 136.3, 134.5, 128.6, 128.5, 128.3, 128.2, 128.1, 127.8, 127.2, 123.1, 121.9, 83.8, 79.6, 71.0, 67.0, 50.9, 38.9, 30.9, 28.2, 28.1. HRMS (ESI) *m/z* calcd for $C_{48}H_{52}N_5O_9$ $(M+H)^+$: 842.3765, found: 842.3759.

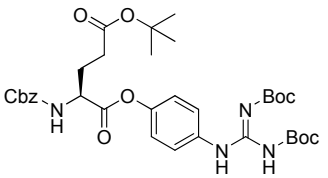
N^{α}-Cbz-L-Asparagine *p*-guanidinophenyl ester (Z-L-Asn-OGp)



General procedure B1 using **3N** (100 mg). The product was obtained as a sticky oil (quant). R_f 0.50 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). $[\alpha]_D^{20}$ –27.0 (*c* 0.38, DMSO).

IR (film) 3329, 3183, 3032, 1761, 1671, 1589, 1507, 1411, 1269, 1200, 1139, 1061 cm^{-1}. ^1H NMR (DMSO-d_6, 300 MHz): δ 9.87 (s, NH), 7.87 (d, *J* = 7.8 Hz, NH), 7.50 (br s, 4NH), 7.40-7.21 (m, 7H), 7.18-7.10 (m, 2H), 7.02 (s, 2NH), 5.08 (s, 2H), 4.64 (dt, *J* = 6.7, 6.7 Hz, 1H), 2.75 (dd, *J* = 6.1, 15.6 Hz, 1H), 2.65 (dd, *J* = 7.0, 15.6 Hz, 1H). ^{13}C NMR (DMSO-d_6, 75 MHz): δ 170.6, 170.3, 155.9, 148.5, 136.7, 132.8, 128.3, 127.8, 127.6, 125.8, 122.6, 65.6, 50.7, 36.7. HRMS (ESI) *m/z* calcd for $C_{19}H_{21}N_5NaO_5$ $(M+Na)^+$: 422.1440, found: 422.1440.

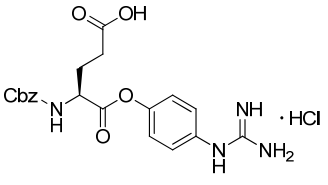
N^{α}-Cbz-L-Glutamate(O^{t}Bu) *p*-[*N',N''*-di(Boc)guanidino]phenyl ester (3E)



General procedure A1 using Z-L-Glu(O^{t}Bu)-OH (472 mg, 1.40 mmol), **2** (351 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:9→1:2) (608 mg, 91%). R_f 0.22 (EtOAc/heptane 1:5). Mp 54 °C. $[\alpha]_D^{20}$ –10.4 (*c* 0.93, CH_2Cl_2). IR

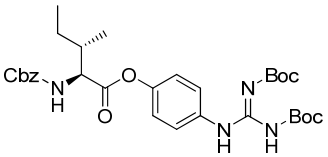
(film) 2976, 2928, 1770, 1719, 1637, 1507, 1411, 1238, 1149, 1113, 1056 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 11.62 (s, NH), 10.34 (s, NH), 7.62 (d, *J* = 8.8 Hz, 2H), 7.39-7.29 (m, 5H), 7.05 (d, *J* = 8.5 Hz, 2H), 5.52 (d, *J* = 8.1 Hz, NH), 5.13 (s, 2H), 4.66-4.55 (m, 1H), 2.52-2.24 (m, 3H), 2.19-2.03 (m, 1H), 1.54 (s, 9H), 1.50 (s, 9H), 1.44 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.9, 170.6, 163.4, 155.9, 153.4, 153.3, 146.9, 136.1, 134.7, 128.5, 128.2, 128.1, 123.1, 121.6, 83.8, 81.0, 79.7, 67.1, 53.7, 31.4, 28.1, 28.0, 27.4. HRMS (ESI) *m/z* calcd for $C_{34}H_{47}N_4O_{10}$ $(M+H)^+$: 671.3292, found: 671.3298.

N$^{\alpha}$-Cbz-L-Glutamate *p*-guanidinophenyl ester (Z-L-Glu-OGp)



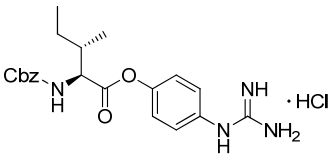
General procedure B1 using **3E** (100 mg). The product was obtained as a sticky oil (quant). R_f 0.50 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). $[\alpha]_D^{20}$ –28.0 (*c* 0.36, MeOH). IR (film) 3321, 3168, 3132, 2950, 1757, 1710, 1671, 1633, 1588, 1507, 1455, 1403, 1261, 1195, 1165, 1059 cm^{-1}. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 10.07 (s, NH), 8.01 (d, *J* = 7.2 Hz, NH), 7.58 (br s, 4NH), 7.40-7.25 (m, 7H), 7.19-7.11 (m, 2H), 5.08 (s, 2H), 4.37-4.27 (m, 1H), 2.42 (t, *J* = 7.3 Hz, 2H), 2.21-2.05 (m, 1H), 2.04-1.87 (m, 1H). ^{13}C NMR (DMSO-d_6, 75 MHz): δ 173.5, 170.8, 156.1, 156.0, 148.3, 136.7, 132.9, 128.3, 127.8, 127.7, 125.7, 122.6, 65.6, 53.3, 31.2, 25.6. HRMS (ESI) *m/z* calcd for $C_{20}H_{23}N_4O_6$ $(M+H)^+$: 415.1618, found: 415.1608.

N$^{\alpha}$-Cbz-L-Isoleucine *p*-[*N'*,*N''*-di(Boc)guanidino]phenyl ester (3I)



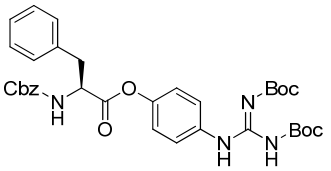
General procedure A1 using Z-L-Ile-OH (372 mg, 1.40 mmol), **2** (351 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:9→1:2) (582 mg, 97%). R_f 0.50 (EtOAc/heptane 1:2). Mp 49 °C. $[\alpha]_D^{20}$ –5.1 (*c* 0.87, CH_2Cl_2). IR (film) 2971, 2924, 2868, 1766, 1717, 1636, 1507, 1410, 1237, 1150, 1113, 1057 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 11.62 (s, NH), 10.34 (s, NH), 7.62 (d, *J* = 8.6 Hz, 2H), 7.42-7.28 (m, 5H), 7.04 (d, *J* = 8.6 Hz, 2H), 5.38 (d, *J* = 8.8 Hz, NH), 5.13 (s, 2H), 4.57 (dd, *J* = 4.7, 8.6 Hz, 1H), 2.13-1.99 (m, 1H), 1.81-1.68 (m, 1H), 1.53 (s, 9H), 1.50 (s, 9H), 1.40-1.20 (m, 1H), 1.05 (d, *J* = 6.8 Hz, 3H), 0.98 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 170.6, 163.4, 156.1, 153.4, 153.2, 146.8, 136.1, 134.7, 128.5, 128.2, 128.1, 123.1, 121.6, 83.8, 79.6, 67.1, 58.4, 38.1, 28.1, 28.0, 25.1, 15.6, 11.6. HRMS (ESI) *m/z* calcd for $C_{31}H_{42}N_4NaO_8$ $(M+Na)^+$: 621.2900, found: 621.2897.

N$^{\alpha}$-Cbz-L-Isoleucine *p*-guanidinophenyl ester (Z-L-Ile-OGp)



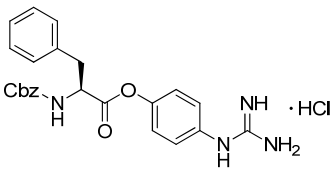
General procedure B1 using **3I** (100 mg). The product was obtained as a sticky oil (quant). R_f 0.46 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). $[\alpha]_D^{20}$ –17.6 (*c* 0.27, MeOH). IR (film) 3312, 3172, 2965, 2928, 1760, 1705, 1673, 1631, 1588, 1507, 1199, 1135 cm^{-1}. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 10.18 (s, NH), 7.95 (d, *J* = 7.3 Hz, NH), 7.63 (br s, 4NH), 7.43-7.25 (m, 7H), 7.18-7.09 (m, 2H), 5.09 (s, 2H), 4.30-4.09 (m, 1H), 2.03-1.87 (m, 1H), 1.63-1.25 (m, 2H), 0.99 (d, *J* = 6.7 Hz, 3H), 0.90 (t, *J* = 7.3 Hz, 3H). ^{13}C NMR (DMSO-d_6, 75 MHz): δ 170.5, 156.4, 156.1, 148.2, 136.8, 132.9, 128.3, 127.8, 127.7, 125.7, 122.6, 65.6, 58.8, 36.0, 24.9, 15.4, 11.1. HRMS (ESI) *m/z* calcd for $C_{21}H_{26}N_4NaO_4$ $(M+Na)^+$: 421.1852, found: 421.1845.

N^{α}-Cbz-L-Phenylalanine *p*-[*N',N''*-di(Boc)guanidino]phenyl ester (3F)



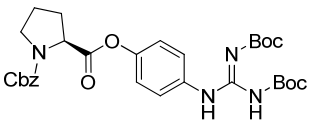
General procedure A1 using Z-L-Phe-OH (419 mg, 1.40 mmol), **2** (351 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:9→1:2) (563 mg, 89%). R_f 0.80 (2% MeOH in CH_2Cl_2). Mp 62 °C. $[\alpha]_D^{20}$ –10.1 (*c* 0.98, CH_2Cl_2). IR (film) 2978, 2933, 1761, 1719, 1640, 1507, 1412, 1239, 1153, 1114, 1057, 620, 607 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.62 (s, NH), 10.34 (s, NH), 7.60 (d, *J* = 8.8 Hz, 2H), 7.39-7.17 (m, 10H), 6.95 (d, *J* = 8.6 Hz, 2H), 5.30 (d, *J* = 8.2 Hz, NH), 5.12 (s, 2H), 4.93-4.83 (m, 1H), 3.25 (d, *J* = 5.8 Hz, 2H), 1.53 (s, 9H), 1.51 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 170.1, 163.4, 155.6, 153.5, 153.3, 146.8, 135.4, 134.7, 129.4, 128.8, 128.5, 128.2, 128.1, 127.4, 123.2, 121.6, 83.9, 79.8, 67.1, 54.9, 38.3, 28.2, 28.1. HRMS (ESI) *m/z* calcd for $C_{34}H_{41}N_4O_8$ $(M+H)^+$: 633.2924, found: 633.2924.

N^{α}-Cbz-L-Phenylalanine *p*-guanidinophenyl ester (Z-L-Phe-OGp)



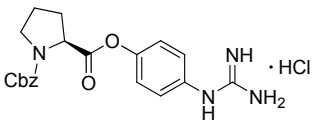
General procedure B1 using **3F** (100 mg). The product was obtained as a sticky oil (quant). R_f 0.44 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). $[\alpha]_D^{20}$ –10.0 (*c* 0.62, MeOH). IR (film) 3330, 3171, 3032, 2954, 1763, 1701, 1672, 1630, 1507, 1198, 1169, 1053 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 9.72 (br, NH), 7.79-6.67 (m, 18H), 5.84-5.66 (m, NH), 5.04 (d, *J* = 12.0 Hz, 1H), 4.97 (d, *J* = 12.4 Hz, 1H), 4.84-4.66 (m, 1H), 3.31-3.01 (m, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.7, 156.7, 155.7, 136.0, 135.5, 135.1, 133.8, 130.0, 129.2, 128.6, 128.5, 128.2, 128.0, 127.2, 125.5, 66.1, 55.0, 38.1. HRMS (ESI) m/z calcd for $C_{24}H_{24}N_4NaO_4$ $(M+Na)^+$: 455.1695, found: 455.1693.

N^{α}-Cbz-L-Proline *p*-[*N',N''*-di(Boc)guanidino]phenyl ester (3P)



General procedure A1 using Z-L-Pro-OH (419 mg, 1.40 mmol), **2** (351 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (MeOH in CH_2Cl_2, 1→2%) (479 mg, 82%). R_f 0.76 (2% MeOH in CH_2Cl_2). Mp 65 °C. $[\alpha]_D^{20}$ –55.4 (*c* 0.72, CH_2Cl_2). IR (film) 2976, 2932, 2876, 1761, 1712, 1636, 1502, 1411, 1238, 1148, 1114, 1057 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.62 (s, NH), 10.32 (s, 0.45NH), 10.31 (s, 0.55NH), 7.60 (d, *J* = 8.8 Hz, 0.9H), 7.52 (d, *J* = 8.9 Hz, 1.1H), 7.44-7.26 (m, 5H), 7.06 (d, *J* = 8.8 Hz, 0.9H), 6.71 (d, *J* = 8.9 Hz, 1.1H), 5.26 (d, *J* = 12.2 Hz, 0.55H), 5.20 (d, *J* = 12.5 Hz, 0.45H), 5.14 (d, *J* = 12.5 Hz, 0.45H), 5.05 (d, *J* = 12.2 Hz, 0.55H),4.58 (dd, *J* = 4.1, 8.5 Hz, 0.45H), 4.52 (dd, *J* = 4.1, 8.5 Hz, 0.55H), 3.74-3.48 (m, 2H), 2.44-1.71 (m, 4H), 1.53 (s, 9H), 1.51 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.2, 171.1, 163.3, 154.8, 154.1, 153.5, 153.2, 147.3, 147.0, 136.6, 136.2, 134.3, 128.5, 128.4, 128.1, 127.9, 127.8, 123.2, 121.7, 121.5, 83.8, 79.7, 67.2, 67.0, 59.3, 58.8, 47.0, 46.4, 31.0, 29.9, 28.1, 28.0, 24.4, 23.6. HRMS (ESI) *m/z* calcd for $C_{30}H_{39}N_4O_8$ $(M+H)^+$: 583.2768, found: 583.2765.

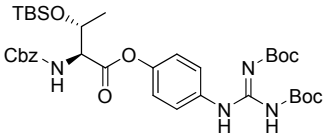
N^{α}-Cbz-L-Proline *p*-guanidinophenyl ester (Z-L-Pro-OGp)



General procedure B1 using **3P** (100 mg). The product was obtained as a sticky oil (quant). R_f 0.45 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). $[\alpha]_D^{20}$ –64.1 (*c* 0.30, MeOH). IR (film) 3331, 3169, 2976, 2950, 2885, 1765, 1697, 1671, 1629, 1507, 1198, 1167, 1142 cm^{-1}. ^1H NMR (CD_3CN, 300 MHz): δ 9.90 (s, 0.5NH), 9.88

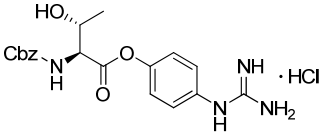
(s, 0.5NH), 7.43-7.32 (m, 5H), 7.30 (d, *J* = 9.0 Hz, 1H), 7.24 (d, *J* = 8.6 Hz, 1H), 7.15 (d, *J* = 8.6 Hz, 1H), 6.99 (br s, 3NH), 6.92 (d, *J* = 8.6 Hz, 1H), 5.18 (d, *J* = 12.4 Hz, 0.5H), 5.16 (d, *J* = 12.4 Hz, 0.5H), 5.11 (d, *J* = 12.6 Hz, 0.5H), 5.04 (d, *J* = 12.4 Hz, 0.5H), 4.52 (dd, *J* = 8.8, 16.0 Hz, 0.5H), 4.51 (dd, *J* = 8.6, 16.6 Hz, 0.5H), 3.59-3.45 (m, 2H), 2.49-1.91 (m, 4H). ^{13}C NMR (CD_3CN, 75 MHz): δ 158.1, 155.7, 154.9, 153.8, 150.7, 150.5, 147.7, 133.0, 129.5, 129.4, 129.1, 128.9, 128.6, 128.1, 128.0, 124.0, 123.9, 67.6, 60.4, 59.8, 47.9, 47.4, 30.8, 30.5, 25.2, 24.3. HRMS (ESI) m/z calcd for $C_{20}H_{23}N_4O_4$ $(M+H)^+$: 383.1719, found: 383.1706.

N^{α}-Cbz-L-Threonine(OTBS) *p*-[*N′,N″*-di(Boc)guanidino]phenyl ester (3T)



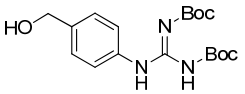
General procedure A1 using Z-L-Thr(OTBS)-OH (515 mg, 1.40 mmol), **2** (351 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:9→1:2) (692 mg, 98%). R_f 0.58 (EtOAc/heptane 1:2). Mp 59 °C. $[\alpha]_D^{20}$ −14.3 (*c* 0.84, CH_2Cl_2). IR (film) 2928, 2854, 1766, 1718, 1638, 1506, 1410, 1237, 1150, 1113, 1058 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 11.63 (s, NH), 10.35 (s, NH), 7.61 (d, *J* = 9.0 Hz, 2H), 7.42-7.31 (s, 5H), 7.08 (d, *J* = 9.0 Hz, 2H), 5.53 (d, *J* = 9.6 Hz, NH), 5.16 (s, 2H), 4.65-4.57 (m, 1H) 4.48 (dd, *J* = 1.7, 9.7 Hz, 1H) 1.54 (s, 9H) 1.51 (s, 9H) 1.29 (d, *J* = 6.2 Hz, 3H) 0.86 (s, 9H) 0.08 (s, 3H), 0.03 (s, 3H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 169.3, 163.5, 156.8, 153.5, 153.3, 147.2, 136.2, 134.5, 128.6, 128.2, 123.0, 121.5, 83.8, 79.8, 68.7, 67.2, 60.2, 28.2, 28.1, 25.7, 21.0, 17.9, -4.2, -4.9. HRMS (ESI) *m/z* calcd for $C_{35}H_{53}N_4O_9Si$ $(M+H)^+$:701.3582, found: 701.3580.

N^{α}-Cbz-L-Threonine *p*-guanidinophenyl ester (Z-L-Thr-OGp)



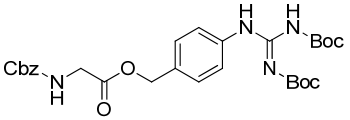
General procedure B1 using **3T** (100 mg). The product was obtained as a sticky oil (quant). R_f 0.45 ($CHCl_3$/MeOH/NH_4OH, 65:45:20). $[\alpha]_D^{20}$ −27.5 (*c* 0.21, MeOH). IR (film) 3330, 3182, 3088, 2980, 1765, 1701, 1671, 1633, 1507, 1196, 1165, 1068 cm^{-1}. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 10.00 (s, NH), 8.65 (br s, NH), 7.64-7.23 (m, 11H), 7.19-7.13 (m, 2H), 5.09 (s, 2H), 4.38-4.19 (m, 2H), 1.20 (d, *J* = 6.3 Hz, 3H). ^{13}C NMR (DMSO-d_6, 75 MHz): δ 169.6, 156.4, 156.0, 148.4, 136.8, 132.8, 128.3, 127.8, 127.6, 125.7, 122.6, 122.5, 66.4, 65.6, 60.3, 19.9. HRMS (ESI) m/z calcd for $C_{19}H_{22}N_4NaO_5$ $(M+Na)^+$: 409.1488, found: 409.1489.

***p*-[*N′, N″*-Di(Boc)guanidino]benzylalcohol (5)**



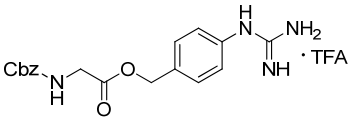
N,N′-Di(Boc)-*S*-methylisothiourea (4.03 g, 13.9 mmol, 1 equiv) and *p*-aminobenzylalcohol (2.56 g, 20.8 mmol, 1.5 equiv) were dissolved in dry THF (80 mL) and this mixture was cooled to 0 °C before $HgCl_2$ (4.14 g, 15.3 mmol, 1.1 equiv) was added. After stirring for 20 min under argon, the temperature was raised to room temperature and the mixture was stirred for 20 h. The reaction mixture was diluted with CH_2Cl_2, filtered and the filtrate was concentrated under reduced pressure. The remaining solid was purified by column chromatography (EtOAc/heptane 1:2) to afford **5** (2.384 g, 65%) as an off-white solid. R_f 0.35 (EtOAc/heptane 1:2). Mp point 25.0 °C. IR (film) 3287, 3261, 2978, 2928, 1719, 1629, 1605, 1563, 1408, 1368, 1337, 1235, 1150, 1118, 1100, 1057, 1028 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 11.63 (s, NH), 10.33 (s, NH), 7.60-7.55 (m, 2H), 7.33-7.28 (m, 2H), 4.63 (s, 2H), 1.53 (s, 9H), 1.50 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 163.1, 153.1, 152.8, 136.9, 135.7, 128.0, 127.1, 121.9, 83.3, 79.2, 64.5, 27.6. HRMS (ESI) *m/z* calcd for $C_{18}H_{28}N_3O_5$ $(M+H)^+$: 366.2029, found: 366.2006.

N^{α}-Cbz-Glycine *p*-[*N',N''*-di(Boc)guanidino]benzyl ester (6G)



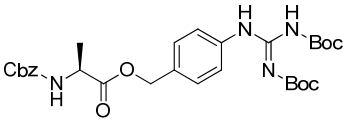
General procedure A2 using Z-Gly-OH (1,47 g, 7.01 mmol), **5** (1.83 g, 5.02 mmol), DMAP (124 mg, 1.02 mmol) and DCC (1.45 g, 7.02 mmol). The product was obtained as an oil after purification by column chromatography (EtOAc/heptane 1:3) (2.72 g, 97%). R_f 0.25 (EtOAc/heptane 1:3). IR (film) 3283, 3257, 2976, 2933, 1718, 1634, 1605, 1558, 1517, 1409, 1368, 1338, 1300, 1236, 1150, 1118, 1100, 1056, 1028 cm^{-1}. ^{1}H-NMR ($CDCl_3$, 400 MHz): δ 11.63 (s, NH), 10.36 (s, NH), 7.63-7.57 (m, 2H), 7.38-7.28 (m, 7H), 5.27(t, *J* = 5.6, NH), 5.13 (ds, 4H), 4.01 (d, *J* = 5.6 Hz, 2H), 1.54 (s, 9H), 1.51 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 169.8, 163.4, 156.2, 153.5, 153.3, 137.1, 136.2, 131.2, 129.2, 128.5, 128.2, 128.1, 122.2, 83.8, 79.7, 67.1, 66.8, 42.8, 28.1. HRMS (ESI) *m/z* calcd for $C_{28}H_{37}N_4O_8$ $(M+H)^+$: 557.2611, found: 557.2607.

N^{α}-Cbz-Glycine *p*-guanidinobenzyl ester (Z-Gly-OGb)



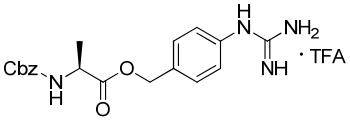
General procedure B2 using **6G** (900 mg, 1.62 mmol). The product was obtained as oily TFA salt (734 mg, quant). IR (film) 3503, 3355, 3179, 1779, 1743, 1725, 1690, 1675, 1640, 1516, 1453, 1409, 1361, 1282, 1256, 1191, 1173, 1051, 1003 cm^{-1}. ^{1}H NMR (CD_3OD, 400 MHz): δ 7.50-7.45 (m, 2H), 7.38-7.29 (m, 5H), 7.29-7.25 (m, 2H), 5.21 (s, 2H), 5.10 (s, 2H), 3.91 (s, 2H). ^{13}C NMR (CD_3OD, 75 MHz): δ 171.7, 161.7, 161.2, 159.2, 137.0, 136.1, 130.8, 129.5, 129.1, 128.8, 126.6, 67.8, 67.0, 43.5. HRMS (ESI) *m/z* calcd for $C_{18}H_{21}N_4O_4$ $(M+H)^+$: 357.1563, found: 357.1557.

N^{α}-Cbz-L-Alanine *p*-[*N',N''*-di(Boc)guanidino]benyl ester (6A)



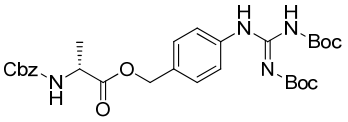
General procedure A2 using Z-L-Ala-OH (295 mg, 1.32 mmol), **5** (345 mg, 0.94 mmol), DMAP (23 mg, 0.19 mmol) and DCC (275 mg, 1.33 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:2) (533 mg, 99%). R_f 0.45 (EtOAc/heptane 1:2). Mp 45.2 °C. $[\alpha]_D^{20}$ −2.97 (*c* 2.12, CH_2Cl_2). IR (film) 3285, 3259, 2979, 2930, 1717, 1635, 1605, 1561, 1517, 1453, 1409, 1367, 1333, 1301, 1235, 1150, 1121, 1100, 1057, 1028 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 11.63 (s, NH), 10.37 (s, NH), 7.63-7.57 (m, 2H), 7.38-7.27 (m, 7H), 5.31 (d, *J* = 7.7 Hz, NH), 5.18-5.06 (m, 2H), 5.11 (s, 2H), 4.42 (dq, *J* = 7.3, 7.4 Hz, 1H), 1.54 (s, 9H), 1.51 (s, 9H), 1.41 (d, *J* = 7.2 Hz, 1H) . ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.7, 163.5, 155.5, 153.4, 153.3, 137.0, 136.3, 131.5, 129.0, 128.5, 128.1, 128.0, 122.1, 83.8, 79.7, 66.9, 66.8, 49.7, 28.2, 28.1, 18.7. HRMS (ESI) *m/z* calcd for $C_{29}H_{39}N_4O_8$ $(M+H)^+$: 571.2768, found: 571.2755.

N^{α}-Cbz-L-Alanine *p*-guanidinobenyl ester (Z-L-Ala-OGb)



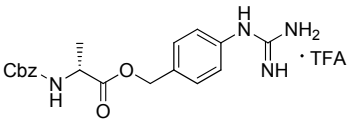
General procedure B2 using **6A** (290 mg, 0.51 mmol). The product was obtained as a sticky oil (quant). $[\alpha]_D^{20}$ −7.8 (*c* 1.28, MeOH). IR (film) 3343, 3174, 1783, 1673, 1610, 1580, 1531, 1517, 1454, 1340, 1299, 1258, 1199, 1174, 1140, 1070 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 9.49 (s, NH), 8.61 (br s, 2NH), 7.46-7.22 (m, 7H), 7.22-7.12 (m, 2H), 6.86 (br s, 3NH), 5.40 (d, *J* = 7.7 Hz, NH), 5.25-5.01 (m, 4H), 4.42-4.30 (m, 1H), 1.43 (d, *J* = 7.2 Hz, 3H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.9, 161.3, 160.7, 156.3, 135.7, 133.6, 129.9, 128.6, 128.4, 127.9, 125.7, 67.4, 66.2, 49.8, 30.9, 18.0. HRMS (ESI) *m/z* calcd for $C_{19}H_{22}N_4O_4$ $(M+H)^+$: 371. 1719, found: 371.1695.

N^{α}-Cbz-D-Alanine p-[*N',N''*-di(Boc)guanidino]benyl ester (6a)



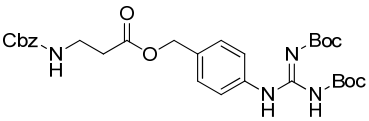
General procedure A2 using Z-D-Ala-OH (313 mg, 1.40 mmol), **5** (365 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:2) (524 mg, 92%). R_f 0.40 (EtOAc/heptane 1:2). Mp 47.9 °C. $[\alpha]_D^{20}$ +3.0 (*c* 2.06, CH_2Cl_2). IR (film) 3281, 3255, 2979, 2937, 1717, 1635, 1606, 1561, 1517, 1410, 1367, 1338, 1301, 1235, 1150, 1121, 1100, 1057, 1028 cm^{-1}. ^1H NMR ($CDCl_3$, 400 MHz): δ 11.63 (s, NH), 10.36 (s, NH), 7.64-7.58 (m, 2H), 7.38-7.27 (m, 7H), 5.32 (d, *J* = 7.5 Hz, NH), 5.18-5.06 (m, 2H) 5.11 (s, 2H), 4.42 (dq, *J* = 7.2, 7.4 Hz, 1H), 1.54 (s, 9H), 1.51 (s, 9H), 1.40 (d, *J* = 7.2 Hz, 3H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.7, 163.4, 155.5, 153.4, 153.3, 137.0, 131.5, 129.0, 128.5, 128.1, 128.0, 122.1, 83.8, 79.7, 66.9, 66.8, 49.7, 28.2, 28.1, 18.7. HRMS (ESI) *m/z* calcd for $C_{29}H_{39}N_4O_8$ $(M+H)^+$: 571.2768, found: 571.2756.

N^{α}-Cbz-D-Alanine *p*-guanidinobenyl ester (Z-D-Ala-OGb)



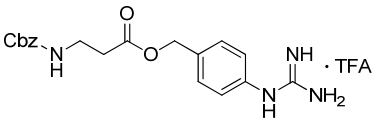
General procedure B2 using **6a** (100 mg). The product was obtained as a sticky oil (quant). $[\alpha]_D^{20}$ +6.1 (*c* 0.86, MeOH). IR (film) 3345, 3188, 1783, 1673, 1601, 1580, 1517, 1454, 1341, 1295, 1256, 1200, 1168, 1139, 1071 cm^{-1}. ^1H NMR (CD_3CN, 400 MHz): δ 9.39 (s, NH), 7.48-7.42 (m, 2H), 7.41-7.29 (m, 5H), 7.29-7.22 (m, 2H), 6.75 (br s, 3NH), 6.03 (d, *J* = 7.6 Hz, NH), 5.16 (s, 2H), 5.06 (s, 2H) 4.26 (dq, *J* = 7.3, 7.4 Hz, 1H), 1.37 (d, *J* = 7.3 Hz, 3H). ^{13}C NMR (CD_3CN, 75 MHz): δ 173.8, 157.7, 136.7, 135.3, 130.3, 129.4, 128.9, 128.7, 126.6, 67.1, 66.6, 50.8, 17.7. HRMS (ESI) *m/z* calcd for $C_{19}H_{22}N_4O_4$ $(M+H)^+$: 371. 1719, found: 371.1709.

N^{α}-Cbz-β-Alanine *p*-[*N',N''*-di(Boc)guanidino]benyl ester (6βA)



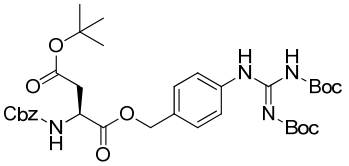
General procedure A2; using Z-β-Ala-OH (313 mg, 1.40 mmol), **5** (365 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:2) (516 mg, 93%). R_f 0.45 (EtOAc/heptane 1:2). Mp 25 °C. IR (film) 3291, 3265, 2978, 2937, 1717, 1634, 1605, 1560, 1517, 1409, 1367, 1338, 1301, 1234, 1149, 1121, 1100, 1057, 1027 cm^{-1}. ^1H NMR ($CDCl_3$, 400 MHz): δ 11.63 (s, NH), 10.35 (s, NH), 7.62-7.57 (m, 2H), 7.38-7.27 (m, 7H), 5.29 (br s, NH), 5.09 (s, 2H), 5.07 (s, 2H), 3.47 (dt, *J* = 6.0, 6.1 Hz, 2H), 2.57 (t, *J* = 6.0 Hz, 2H), 1.54 (s, 9H), 1.50 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.0, 163.4, 156.2, 153.4, 153.2, 136.9, 136.4, 131.8, 129.0, 128.5, 128.1, 128.0, 122.2, 83.8, 79.7, 66.6, 66.1, 36.5, 34.4, 28.1, 28.0. HRMS (ESI) *m/z* calcd for $C_{29}H_{39}N_4O_8$ $(M+H)^+$: 571.2768, found: 571.2789.

N^{α}-Cbz-β-Alanine *p*-guanidinobenyl ester (Z-β-Ala-OGb)



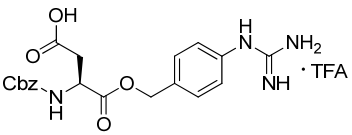
General procedure B2 using **6βA** (100 mg). The product was obtained as a sticky oil (quant). IR (film) 3330, 3175, 3062, 2946, 1754, 1701, 1672, 1629, 1588, 1508, 1455, 1257, 1200, 1166, 1065, 1014 cm^{-1}. ^1H NMR (CD_3CN, 400 MHz): δ 9.46 (s, NH), 7.49-7.42 (m, 2H), 7.39-7.21 (m, 7H), 6.79 (br s, 4NH), 5.74 (br s, NH), 5.11 (s, 2H), 5.03 (s, 2H), 3.37 (dt, *J* = 6.3, 6.5 Hz, 2H), 2.55 (t, *J* = 6.6 Hz, 2H). ^{13}C NMR (CD_3CN, 75 MHz): δ 171.4, 158.1, 157.3, 150.8, 132.8, 129.4, 128.8, 128.6, 128.0, 124.3, 66.9, 37.5, 35.5. HRMS (ESI) *m/z* calcd for $C_{19}H_{23}N_4O_4$ $(M+H)^+$: 371.1719, found: 371.1700.

N^{α}-Cbz-L-Aspartate(O^{t}Bu) *p*-[*N',N''*-di(Boc)guanidino]benyl ester (6D)



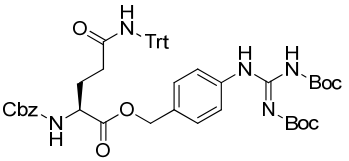
General procedure A2 using Z-L-Asn(O^{t}Bu)-OH (453 mg, 1.40 mmol), **5** (365 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:2) (599 mg, 89%). R_f 0.55 (EtOAc/heptane 1:2). Mp 46.6 °C. $[\alpha]_D^{20}$ +7.3 (*c* 2.08, CH_2Cl_2). IR (film) 3447, 3417, 3287, 3252, 2977, 2932, 1719, 1636, 1606, 1561, 1504, 1454, 1410, 1367, 1337, 1300, 1234, 1149, 1122, 1100, 1057, 1027 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 11.63 (s, NH), 10.35 (s, NH), 7.62-7.57 (m, 2H), 7.38-7.25 (m, 7H), 5.75 (d, *J* = 8.6 Hz, NH), 5.18-5.08 (m, 2H), 5.12 (s, 2H), 4.61 (dt, *J* = 4.5, 8.7, 1H), 2.94 (dd, *J* = 4.6, 17.0, 1H), 2.75 (dd, *J* = 4.5, 17.0, 1H), 1.54 (s, 9H), 1.50 (s, 9H), 1.40 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 170.7, 169.9, 163.5, 156.0, 153.4, 153.3, 137.0, 131.4, 129.0, 128.5, 128.1, 128.0, 122.1, 83.8, 81.8, 79.7, 67.1, 50.6, 37.7, 28.2, 28.1, 28.0. HRMS (ESI) *m/z* calcd for $C_{34}H_{47}N_4O_{10}$ $(M+H)^+$: 671.3292, found: 671.3251.

N^{α}-Cbz-L-Aspartate *p*-guanidinobenyl ester (Z-L-Asp-OGb)



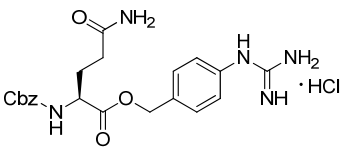
General procedure B2 using **6D** (150 mg). The product was obtained as a sticky oil (quant). $[\alpha]_D^{20}$ –5.7 (*c* 0.38, MeOH). IR (film) 3331, 3179, 3032, 2935, 2850, 1689, 1580, 1517, 1454, 1411, 1368, 1339, 1260, 1200, 1140, 1064 cm^{-1}. ^{1}H NMR (CD_3CN, 400 MHz): δ 9.46 (s, NH), 7.46-7.40 (m, 2H), 7.40-7.29 (m, 5H), 7.27-7.22 (m, 2H), 6.78 (br s, 3NH), 6.15 (d, *J* = 8.7, NH), 5.16 (s, 2H), 5.10 (d, *J* = 12.6 Hz, 1H), 5.06 (d, *J* = 12.5 Hz, 1H), 4.58 (dt, *J* = 5.8, 8.6 Hz, 1H), 2.90-2.79 (m, 2H). ^{13}C NMR (CD_3CN, 75 MHz): δ 172.2, 171.8, 157.7, 136.4, 135.3, 130.3, 129.4, 128.9, 128.7, 126.6, 67.3, 67.0, 51.6, 36.5. HRMS (ESI) *m/z* calcd for $C_{20}H_{23}N_4O_6$ $(M+H)^+$: 415.1618, found: 415.1615.

N^{α}-Cbz-L-Glutamine(Trt) *p*-[*N',N''*-di(Boc)guanidino]benyl ester (6Q)



General procedure A2 using Z-L-Glu(Trt)-OH (731 mg, 1.40 mmol), **5** (365 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:2) (864 mg, 99%). R_f 0.352 (EtOAc/heptane 1:2). Mp 157.2 °C. $[\alpha]_D^{20}$ +3.4 (*c* 1.96, CH_2Cl_2). IR (film) 3408, 3264, 3062, 3032, 2949, 2924, 1720, 1638, 1606, 1535, 1492, 1446, 1410, 1365, 1338, 1235, 1149, 1122, 1100, 1057 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 11.62 (s, NH), 10.34 (s, NH), 7.59-7.52 (m, 2H), 7.38-7.13 (m, 22H), 6.78 (br s, NH), 5.57 (d, *J* = 7.6 Hz, NH), 5.14-5.02 (m, 2H), 5.09 (s, 2H), 4.41-4.31 (m, 1H), 2.46-2.24 (m, 2H), 2.04-1.88 (m, 2H), 1.54 (s, 9H), 1.50 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.7, 170.7, 163.4, 156.3, 153.4, 144.5, 137.1, 136.2, 131.4, 129.2, 128.7, 128.5, 128.1, 128.0, 83.8, 79.7, 67.0, 66.9, 53.7, 35.4, 33.3, 28.1. HRMS (ESI) *m/z* calcd for $C_{50}H_{56}N_5O_9$ $(M+H)^+$: 870.4078, found: 870.4043.

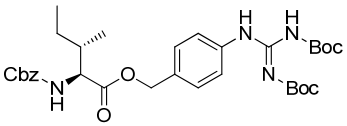
N^{α}-Cbz-L-Glutamine *p*-guanidinobenyl ester (Z-L-Gln-OGb)



General procedure B1 using **6Q** (200 mg, 0.23 mmol) in the presence of triethylsilane (37 μL, 0.23 mmol). The product was obtained as a sticky oil (90 mg, 82%). $[\alpha]_D^{20}$ –6.3 (*c* 0.99, MeOH). IR (film) 3330, 3067, 3028, 2954, 1702, 1666, 1531, 1453, 1413, 1341, 1215, 1056 cm^{-1}. ^{1}H NMR (DMSO-d_6, 400

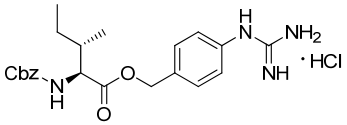
MHz): δ 10.09 (s, NH), 7.81 (d, *J* = 7.7 Hz, NH), 7.58 (br s, 4NH), 7.45-7.17 (m, 9H), 6.79 (br s, 2NH), 5.18-4.99 (m, 2H), 4.13-4.05 (m, 1H), 2.16 (t, *J* = 7.4 Hz, 2H), 2.05-1.93 (m, 1H), 1.82-1.71 (m, 1H). ^{13}C NMR (DMSO-d_6, 75 MHz): δ 173.2, 172.1, 156.1, 155.9, 136.8, 135.1, 133.8, 129.1, 128.3, 127.8, 127.7, 124.1, 65.5, 65.3, 53.5, 30.9, 26.2. HRMS (ESI) *m/z* calcd for $C_{21}H_{26}N_5O_5$ $(M+H)^+$: 428.1934, found: 428.1922.

N^{α}-Cbz-L-Isoleucine *p*-[*N',N''*-di(Boc)guanidino]benyl ester (6I)



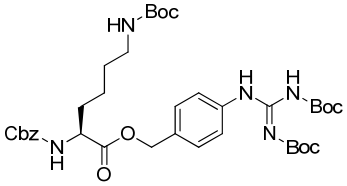
General procedure A2 using Z-L-Ile-OH (372 mg, 1.40 mmol), **5** (365 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:2) (577 mg, 94%). R_f 0.55 (EtOAc/heptane 1:2). Mp 46.0 °C. $[\alpha]_D^{20}$ +5.1 (*c* 2.02, CH_2Cl_2). IR (film) 3432, 3287, 3253, 2974, 2928, 1717, 1636, 1606, 1561, 1515, 1454, 1410, 1367, 1337, 1301, 1233, 1150, 1122, 1099, 1057, 1028 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 11.63 (s, NH), 10.37 (s, NH), 7.64-7.58 (m, 2H), 7.40-7.27 (m, 7H), 5.27 (d, *J* = 8.8 Hz, NH), 5.15 (d, *J* = 12.2 Hz, 1H), 5.11 (s, 2H), 5.08 (d, *J* = 12.2 Hz, 1H), 4.37 (dd, *J* = 4.7, 9.0 Hz, 1H), 1.93-1.84 (m, 1H), 1.54 (s, 9H), 1.51 (s, 9H), 1.44-1.29 (m, 1H), 1.21-1.07 (m, 1H), 0.94-0.84 (m, 6H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.8, 163.5, 156.1, 153.4, 153.3, 137.1, 136.3, 131.5, 129.2, 128.5, 128.3, 128.1, 128.0, 122.1, 83.8, 79.7, 67.0, 66.6, 58.4, 38.1, 28.2, 28.1, 24.9, 15.5, 11.6. HRMS (ESI) *m/z* calcd for $C_{32}H_{45}N_4NaO_8$ $(M+H)^+$: 613.3237, found: 613.3209.

N^{α}-Cbz-L-Isoleucine *p*-guanidinobenyl ester (Z-L-Ile-OGb)



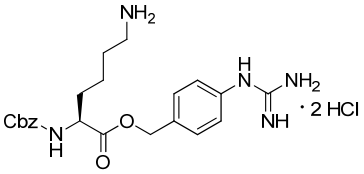
General procedure B1 using **6I** (270 mg). The product was obtained as a sticky oil (331 mg, quant). $[\alpha]_D^{20}$ +0.9 (*c* 2.10, MeOH). IR (film) 3347, 3179, 2967, 2932, 2889, 1779, 1674, 1606, 1581, 1517, 1455, 1260, 1338, 1200, 1157, 1091, 1042 cm^{-1}. ^{1}H NMR (CD_3CN, 400 MHz): δ 9.05 (s, NH), 7.49-7.43 (m, 2H), 7.40-7.29 (m, 5H), 7.28-7.22 (m, 2H), 6.62 (br s, 3NH), 5.98 (d, *J* = 8.6 Hz, NH), 5.17 (s, 2H), 5.07 (s, 2H), 4.18 (dd *J* = 5.9, 8.6 Hz, 1H), 1.91-1.81 (m, 1H), 1.48-1.37 (m, 1H), 1.26-1.11 (m, 1H), 0.90 (d, *J* = 6.9 Hz, 3H), 0.87 (t, *J* = 7.4 Hz, 3H). ^{13}C NMR (CD_3CN, 75 MHz): δ 172.7, 160.5, 160.0, 157.6, 136.7, 135.1, 130.5, 129.4, 128.9, 128.6, 126.7, 67.2, 66.5, 59.9, 37.9, 25.8, 15.9, 11.6. HRMS (ESI) *m/z* calcd for $C_{22}H_{29}N_4O_4$ $(M+H)^+$: 413.2189, found: 413.2197.

N^{α}-Cbz-L-Lysine(Boc) *p*-[*N',N''*-di(Boc)guanidino]benyl ester (6K)



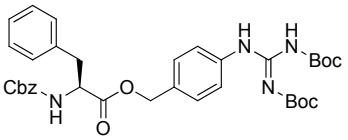
General procedure A2 using Z-L-Lys(Boc)-OH (1028 mg, 2.70 mmol), **5** (703 mg, 1.92 mmol), DMAP (48 mg, 0.39 mmol) and DCC (558 mg, 2.70 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:3) (1.26 g, 90%). R_f 0.60 (EtOAc/heptane 1:2). Mp 50.7 °C. $[\alpha]_D^{20}$ −3.8 (*c* 1.96, CH_2Cl_2). IR (film) 3291, 3257, 2975, 2928, 2860, 2245, 1714, 1631, 1605, 1516, 1410, 1366, 1338, 1301, 1238, 1152, 1057, 1121, 1100, 1057 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 11.62 (s, NH), 10.36 (s, NH), 7.62-7.56 (m, 2H), 7.38-7.27 (m, 7H), 5.37 (d, *J* = 7.5 Hz, NH), 5.15 (d, *J* = 12.2 Hz, 1H), 5.10 (s, 2H), 5.07 (d, *J* = 12.0 Hz, 1H), 4.58 (br s, NH), 4.42-4.34 (m, 1H), 3.12-2.98 (m, 2H), 1.89-1.75 (m, 2H), 1.73-1.59 (m, 2H), 1.54 (s, 9H), 1.50 (s, 9H), 1.42 (s, 9H), 1.39-1.21 (m, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.2, 163.5, 156.0, 153.5, 153.3, 137.1, 131.5, 129.2, 128.5, 128.1, 122.2, 83.8, 79.7, 66.8, 53.8, 40.0, 32.2, 29.5, 28.4, 28.2, 28.1, 22.3. HRMS (ESI) *m/z* calcd for $C_{37}H_{54}N_5O_{10}$ $(M+H)^+$: 728.3871, found: 728.3856.

N^{α}-Cbz-L-Lysine *p*-guanidinobenyl ester (Z-L-Lys-OGb)



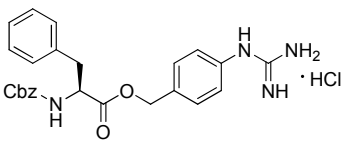
General procedure B1 using **6K** (500 mg, 0.69 mmol). The product was obtained as a sticky oil (221 mg, 65%). $[\alpha]_D^{20}$ –5.5 (*c* 1.32, MeOH). IR (film) 3321, 3144, 3032, 2950, 1705, 1671, 1631, 1605, 1579, 1516, 1454, 1395, 1338, 1256, 1202, 1170, 1135, 1049, 1025 cm^{-1}. ^1H NMR (CD_3OD, 400 MHz): δ 7.49-7.44 (m, 2H), 7.37-7.24 (m, 7H), 5.23 (d, *J* = 12.7 Hz, 1H) 5.18 (d, *J* = 12.6 Hz, 1H), 5.11 (d, *J* = 12.6 Hz, 1H), 5.07 (d, *J* = 12.4 Hz, 1H), 4.22 (dd, *J* = 4.8, 9.6 Hz, 1H), 2.96-2.83 (m, 2H), 1.95-1.84 (m, 1H), 1.79-1.59 (m, 3H), 1.57-1.41 (m, 2H). ^{13}C NMR (CD_3OD, 75 MHz): δ 173.7, 158.8, 136.9, 136.1, 130.8, 129.5, 129.1, 128.9, 126.5, 67.7, 67.2, 55.4, 40.6, 32.0, 28.0, 24.0. HRMS (ESI) *m/z* calcd for $C_{22}H_{30}N_5O_4$ $(M+H)^+$: 428.2298, found: 428.2286.

N^{α}-Cbz-L-Phenylalanine *p*-[*N',N''*-di(Boc)guanidino]benyl ester (6F)



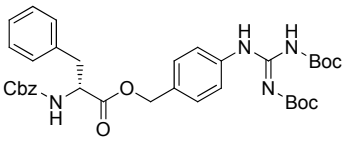
General procedure A1 using Z-L-Phe-OH (419 mg, 1.40 mmol), **5** (365 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:5→1:2) (352 mg, 54%). R_f 0.20 (EtOAc/heptane 1:5). $[\alpha]_D^{20}$ +1.4 (*c* 1.97, CH_2Cl_2). IR (film) 3417, 3261, 2979, 2928, 1711, 1640, 1625, 1603, 1505, 1409, 1392, 1304, 1236, 1150, 1056 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.64 (s, NH), 10.36 (s, NH), 7.65-7.56 (m, 2H), 7.41-7.17 (m, 10H), 7.07-6.98 (m, 2H), 5.26 (d, *J* = 7.9 Hz, NH), 5.15-5.01 (m, 2H), 5.08 (s, 2H), 4.76-4.63 (m, 1H), 3.17-3.01 (m, 2H), 1.54 (s, 9H), 1.51 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.2, 163.4, 155.5, 153.4, 153.2, 137.0, 136.2, 135.5, 131.2, 129.3, 129.2, 128.5, 128.4, 128.1, 128.0, 122.1, 83.7, 79.6, 66.9, 66.8, 54.8, 38.1, 28.1, 28.0. HRMS (ESI) *m/z* calcd for $C_{35}H_{43}N_4O_8$ $(M+H)^+$: 647.3081, found: 647.3075.

N^{α}-Cbz-L-Phenylalanine *p*-guanidinobenyl ester (Z-L-Phe-OGb)



General procedure B1 using **6F** (103 mg). The product was obtained as a sticky oil (110 mg, quant). $[\alpha]_D^{20}$ –1.8 (*c* 1.01, MeOH). IR (film) 3417, 3365, 3257, 2980, 1710, 1625, 1602, 1506, 1409, 1341, 1304, 1234, 1151, 1121, 1055 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 9.90 (br s, NH), 7.38-7.01 (m, 14H), 5.37 (d, *J* = 7.5 Hz, NH), 5.17-4.94 (m, 4H), 4.69-4.57 (m, 1H), 3.15-3.01 (m, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 173.3, 158.4, 158.1, 138.2, 136.8, 130.8, 130.2, 129.6, 129.5, 129.0, 128.7, 127.9, 126.5, 67.6, 67.2, 57.2, 38.6. HRMS (ESI) *m/z* calcd for $C_{25}H_{27}N_4O_4$ $(M+H)^+$: 447.2032, found: 447.2049.

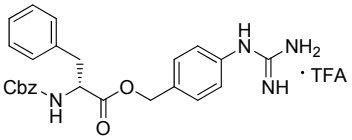
N^{α}-Cbz-D-Phenylalanine *p*-[*N',N''*-di(Boc)guanidino]benyl ester (6f)



General procedure A2 using Z-D-Phe-OH (421 mg, 1.41 mmol), **5** (365 mg, 1.00 mmol), DMAP (28 mg, 0.23 mmol) and DCC (290 mg, 1.41 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:3) (567 mg, 88%). R_f 0.60 (EtOAc/heptane 1:2). Mp 132.2 °C. $[\alpha]_D^{20}$ –2.4 (*c* 2.07, CH_2Cl_2). IR (film) 3278, 3265, 2971, 2928, 1717, 1633, 1605, 1563, 1515, 1411, 1368, 1340, 1301, 1237, 1151, 1122, 1100, 1057, 1028 cm^{-1}. ^1H NMR ($CDCl_3$, 400 MHz): δ 11.63 (s, NH), 10.37 (s, NH), 7.63-7.57 (m, 2H), 7.39-7.29 (m, 5H), 7.25-7.20 (m, 5H), 7.06-6.99 (m, 2H), 5.21 (d, *J* = 7.9 Hz, NH), 5.14-5.01 (m, 2H), 5.10 (s, 2H), 4.69 (dt, *J* = 5.9, 14.0 Hz, 1H), 3.16-3.03 (m,

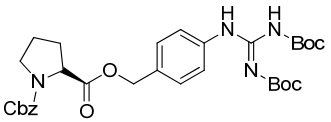
2H), 1.54 (s, 9H), 1.51 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.3, 163.5, 155.6, 153.5, 153.3, 137.1, 136.2, 135.5, 131.2, 129.3, 128.6, 128.5, 128.2, 128.1, 127.1, 122.2, 83.8, 79.7, 66.9, 66.8, 54.8, 38.2, 28.2, 28.1. HRMS (ESI) *m/z* calcd for $C_{35}H_{43}N_4O_8$ $(M+H)^+$: 647.3081, found: 647.3064.

N^{α}-Cbz-D-Phenylalanine *p*-guanidinobenyl ester (Z-D-Phe-OGb)



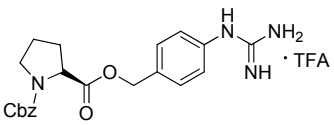
General procedure B2 using **6f** (175 mg). The product was obtained as a sticky oil (150 mg, quant). $[\alpha]_D^{20}$ –1.5 (*c* 1.45, MeOH). IR (film) 3322, 3179, 3032, 2937, 2855, 1673, 1602, 1517, 1453, 1429, 1344, 1259, 1200, 1182, 1082, 1056, 1022 cm^{-1}. ^{1}H NMR (CD_3OD, 400 MHz): δ 7.41-7.36 (m, 2H), 7.35-7.16 (m, 12H), 5.18 (d, *J* = 12.5 Hz, 1H), 5.12 (d, *J* = 12.6 Hz, 1H), 5.04 (s, 2H), 4.49-4.43 (m, 1H), 3.14 (dd, *J* = 6.0, 13.9 Hz, 1H), 2.97 (dd, *J* = 9.0, 13.8 Hz, 1H) . ^{13}C NMR (CD_3OD, 75 MHz): δ 173.3, 161.5, 161.0, 138.2, 136.8, 136.1, 130.8, 130.3, 129.6, 129.5, 129.0, 128.7, 127.9, 126.5, 67.6, 67.2, 57.2, 38.6. HRMS (ESI) *m/z* calcd for $C_{25}H_{27}N_4O_4$ $(M+H)^+$: 447.2032, found: 447.2017.

N^{α}-Cbz-L-Proline *p*-[*N',N''*-di(Boc)guanidino]benyl ester (6P)



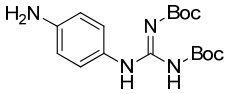
General procedure A1 using Z-L-Pro-OH (349 mg, 1.40 mmol), **5** (365 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). The product was obtained as a white solid after purification by column chromatography (MeOH in CH_2Cl_2, 1→2.5%) (495 mg, 83%). R_f 0.45 (5% MeOH in CH_2Cl_2). $[\alpha]_D^{20}$ –30.8 (*c* 2.00, CH_2Cl_2). IR (film) 3257, 2977, 2872, 1710, 1634, 1605, 1410, 1341, 1301, 1235, 1150, 1120, 1100, 1057 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 11.63 (s, NH), 10.36 (s, 0.5NH), 10.34 (s, 0.5NH), 7.63-7.50 (m, 2H), 7.42-7.24 (m, 6H), 7.21-7.13 (m, 1H), 5.24-4.89 (m, 4H), 4.43 (dd, *J* = 3.4, 8.6 Hz, 0.5H), 4.32 (dd, *J* = 3.4, 8.6 Hz, 0.5H), 3.69-3.41 (m, 2H), 2.30-2.11 (m, 1H), 2.05-1.80 (m, 3H), 1.54 (s, 9H), 1.51 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.5, 172.3, 163.3, 154.8, 154.2, 153.4, 153.2, 136.7, 136.6, 136.5, 132.0, 131.8, 128.8, 128.3, 127.8, 127.7, 127.7, 122.1, 83.7, 79.7, 66.9, 66.3, 66.2, 59.2, 58.8, 46.8, 46.3, 30.8, 29.7, 28.1, 28.0, 24.2, 23.4. HRMS (ESI) *m/z* calcd for $C_{31}H_{41}N_4O_8$ $(M+H)^+$: 597.2924, found: 597.2924.

N^{α}-Cbz-L-Proline *p*-guanidinobenyl ester (Z-L-Pro-OGb)



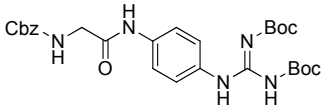
General procedure B2 using **6P** (103 mg). The product was obtained as a sticky oil (108 mg, quant). $[\alpha]_D^{20}$ –24.0 (*c* 1.06, MeOH). IR (film) 3329, 3166, 2967, 2281, 1745, 1680, 1597, 1576, 1517, 1425, 1354, 1273, 1201, 1174, 1133, 1089 cm^{-1}. ^{1}H NMR (CD_3OD, 400 MHz): δ 7.48-7.45 (m, 1H), 7.38-7.24 (m, 7H), 7.23-7.18 (m, 1H), 5.23 (d, *J* = 12.6 Hz, 0.5H), 5.18 (d, *J* = 12.3 Hz, 0.5H), 5.13 (d, *J* = 2.2 Hz, 1H), 5.10 (d, *J* = 12.6 Hz, 0.5H), 5.08 (d, *J* = 12.5 Hz, 0.5H), 5.04 (d, *J* = 12.7 Hz, 0.5H), 5.03 (d, *J* = 12.5 Hz, 0.5H), 4.43 (dd, *J* = 3.9, 8.8 Hz, 0.5H), 4.39 (dd, *J* = 3.7, 8.7 Hz, 0.5H), 3.62-3.46 (m, 2H), 2.37-2.26 (m, 1H), 2.06-1.88 (m, 3H). ^{13}C NMR (CD_3OD, 75 MHz): δ 174.2, 174.0, 138.1, 137.9, 137.0, 136.7, 136.1, 130.7, 129.6, 129.2, 128.8, 126.6, 126.4, 68.3, 67.1, 60.8, 60.4, 31.8, 30.9, 25.3, 24.5, 19.2, 18.4. HRMS (ESI) m/z calcd for $C_{21}H_{25}N_4O_4$ $(M+H)^+$: 397.1876, found: 397.1856.

_p_-[_N'_, _N''_-Di(Boc)guanidino]aniline (8)



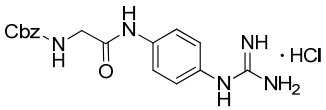
_N,N'_-di(Boc)-_S_-methylisothiourea (2.90 g, 10.0 mmol, 1 equiv) and 1,4-diaminobenzene (3.24 g, 30.0 mmol, 3 equiv) were dissolved in dry THF (50 mL) and this mixture was cooled to 0 °C before $HgCl_2$ (2.99 g, 11.0 mmol, 1.1 equiv) was added. After stirring for 20 min under argon, the temperature was raised to room temperature and the mixture was stirred for 20 h. The white precipitate that was formed during the reaction was filtered off and the filtrate was concentrated under reduced pressure. The remaining solid was purified by column chromatography (EtOAc/heptane 1:2) to afford **8** (1.93 g, 55%) as an off-white solid. R_f 0.31(EtOAc/heptane 1:2). Mp dec at > 300 °C. IR (film) 3430, 3339, 3257, 2976, 1717, 1635, 1621, 1601, 1408, 1332, 1151, 1106, 1061, 610 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.63 (s, NH), 10.09 (s, NH), 7.37-7.31 (m, 2H), 6.67-6.61 (m, 2H), 3.59 (br s, 2NH), 1.53 (s, 9H), 1.49 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 163.7, 153.6, 143.7, 128.0, 124.0, 115.3, 83.4, 79.3, 28.2, 28.1. HRMS (ESI) *m/z* calcd for $C_{17}H_{27}N_4O_4$ $(M+H)^+$: 351.2032, found: 351.2039. Spectral data were in accordance with those reported in literature.[10]

_N_α-Cbz-Glycine _p_-[_N',N''_-di(Boc)guanidino]anilide (9G)



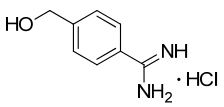
General procedure A1 using Z-Gly-OH (146 mg, 0.70 mmol), **8** (175 mg, 0.50 mmol), DMAP (12 mg, 0.1 mmol) and DCC (144 mg, 0.70 mmol). The product was obtained as a white solid after purification by column chromatography (2% MeOH in CH_2Cl_2) (282 mg, quant). R_f 0.27 (2% MeOH in CH_2Cl_2). Mp 121 °C. IR (film) 3283, 2980, 2928, 1718, 1640, 1515, 1409, 1232, 1152, 1057 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.61 (s, NH), 10.26 (s, NH), 7.97 (br s, NH), 7.52-7.45 (m, 2H), 7.44-7.32 (m, 7H), 5.45 (br s, NH), 5.17 (s, 2H), 3.96 (d, *J* = 5.8 Hz, 2H), 1.54 (s, 9H), 1.48 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 166.7, 163.3, 154.6, 153.2, 136.1, 135.1, 132.4, 128.6, 128.3, 128.1, 124.0, 120.3, 83.9, 79.8, 67.3, 45.1, 28.2, 28.1. HRMS (ESI) *m/z* calcd for $C_{27}H_{35}N_5NaO_7$ $(M+Na)^+$: 564.2434, found: 564.2441.

_N_α-Cbz-Glycine _p_-guanidino anilide (Z-Gly-NGp)



General procedure B1 using **9G** (100 mg, 0.185 mmol). The product was obtained as an amorphous solid (69.8 mg, quant). IR (film) 3365, 2920, 2846, 1668, 1614, 1515, 1455, 1260, 1053, 699 cm^{-1}. ^1H NMR (CD_3OD, 300 MHz): δ 7.72-7.62 (m, 2H), 7.42-7.20 (m, 7H), 5.12 (s, 2H), 3.95 (s, 2H). ^{13}C NMR (CD_3OD, 75 MHz): δ 170.6, 158.3, 139.3, 131.6, 129.5, 129.1, 128.9, 127.5, 122.6, 67.9, 45.4. HRMS (ESI) *m/z* calcd for $C_{17}H_{20}N_5O_3$ $(M+H)^+$: 342.1566, found: 342.1573.

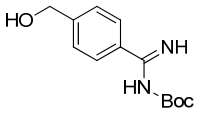
_p_-Amidinobenzyl alcohol·HCl (12)



Hydrochloric acid (g) was dried by guiding it through two washing bottles of conc. H_2SO_4. A cooled (0 °C) solution of *p*-(hydroxymethyl)benzonitrile (2.0 g, 15 mmol, 1 equiv) in abs. EtOH (10 mL) was saturated with dry HCl (g). The resulting reaction mixture was kept at –20 °C for 48 hours to induce crystallization of the imidate·HCl salt. The clear crystals were filtered and immediately 10% NH_3 in abs. EtOH (10 mL) was added in portions and stirred during 24 hours. The following 48 hours ammoniumchloride precipitated at room temperature. After filtration and *in vacuo* evaporation, the crude product was dissolved in water and acidified with conc. HCl (aq). The pure product crystallized as a white solid, was filtered and further dried under high vacuum to afford **12**·HCl (2.3 g, 80%). R_f 0.0 (5% MeOH in CH_2Cl_2). IR (film) 3342, 3135, 3052, 2858, 1667, 1612, 1480, 1397, 1210, 1162, 1051, 1009, 815, 705, 532 cm^{-1}. ^1H NMR (CD_3OD, 300 MHz): δ 7.80 (d, *J* = 10.5 Hz, 2H), 7.59 (d, *J* = 8.7 Hz, 2H), 4.73 (s, 2H). ^{13}C NMR (CD_3OD, 75 MHz): δ 150.1, 129.0, 128.3, 64.2.

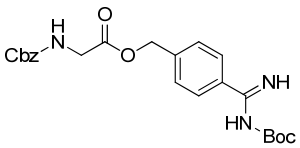
HRMS (ESI) *m/z* calcd for $C_8H_{11}N_2O$ $(M+H)^+$: 151.0871, found: 151.0871. Spectral data were in accordance with those reported in literature.[11]

_p_-[_N'_-Boc-amidino]benzyl alcohol (13)



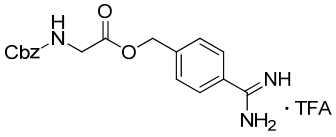
To a solution **12·HCl** (111 mg, 0.6 mmol, 1.0 equiv) in MeOH (30 mL) and TEA (2 mL) was added di-^{t}butyl-dicarbonate (705 mg, 3.2 mmol, 5.4 equiv) and the mixture was refluxed for 5 hours. The reaction mixture was evaporated to dryness and purified with column chromatography (EtOAc/heptane 1:1→3:1) to afford **13** (96 mg, 65%) as a white solid. R_f 0.72 (EtOAc/heptane 3:1). IR (film) 3340, 3129, 3052, 2858, 1667, 1612, 1458, 1320, 1238, 1162, 1044, 870, 746, 536 cm^{-1}. ^{1}H NMR (CD_3OD, 300 MHz): δ 7.79 (d, *J* = 8.4 Hz, 2H), 7.44 (d, *J* = 8.7 Hz, 2H), 4.67 (s, 2H), 1.53 (s, 9H). ^{13}C NMR (CD_3OD, 75 MHz): δ 209.9, 149.8, 128.9, 127.8, 64.6, 28.5. HRMS (ESI) *m/z* calcd for $C_{13}H_{19}N_2O_3$ $(M+H)^+$: 251.1396, found: 251.393.

_N_$^\alpha$-Cbz-Glycine _p_-[_N'_-Boc-amidino]benzyl ester (14)



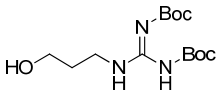
General procedure A1 with adjusted amounts of every component using **13** (96 mg, 0.4 mmol, 1.0 equiv), DCC (106 mg, 0.5 mmol, 1.2 equiv), DMAP (56 mg, 0.5 mmol, 1.2 equiv) and Z-Gly-OH (100 mg, 0.5 mmol, 1.2 equiv). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:1) (162 mg, 95%). R_f 0.28 (EtOAc/heptane 1:1). IR (film) 3391, 2975, 2498, 1715, 1612, 1528, 1494, 1452, 1355, 1286, 1238, 1162, 1134, 1044, 988, 878, 836, 746, 691 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 7.79 (d, *J* = 8.1 Hz, 2H), 7.27-7.39 (m, 7H), 5.47 (t, *J* = 5.7 Hz, NH), 5.17 (s, 2H), 5.10 (s, 2H), 3.98 (d, *J* = 5.7 Hz, 2H), 1.54 (s, 9H). ^{13}C NMR (CD_3OD, 75 MHz): δ 171.7, 159.1, 147.2, 141.3, 138.1, 136.0, 129.5, 129.3, 129.0, 128.9, 128.8, 127.7, 80.5, 67.8, 67.0, 43.5. The product was contaminated with ~ 10% DCU according to NMR. HRMS (ESI) *m/z* calcd for $C_{23}H_{28}N_3O_6$ $(M+H)^+$: 442.1978, found: 442.1979.

_N_$^\alpha$-Cbz-glycine _p_-amidinobenzyl ester (Z-Gly-OAb)



To a suspension of **14** (89 mg, 0.2 mmol) in CH_2Cl_2 (1 mL) was added TFA (1 mL) resulting in a clear solution. After stirring for 2 hours the solvent was evaporated *in vacuo*, the residue was co-evaporated with ^{t}BuOH (3 × 5 mL) and additionally dried under high vacuum to afford **Z-Gly-OAb·TFA** (67 mg, 73%) as a white solid. R_f 0.0 (5% MeOH in CH_2Cl_2). IR (film) 3322, 2920, 2851, 1646, 1563, 1522, 1425, 1314, 1231, 1168, 1085, 885, 608, 428 cm^{-1}. ^{1}H NMR (CD_3OD, 300 MHz): δ 7.78 (d, *J* = 8.4 Hz, 2H), 7.60 (d, *J* = 8.1 Hz, 2H), 7.35-7.31 (m, 5H), 5.30 (s, 2H), 5.11 (s, 2H), 3.96 (s, 2H). ^{13}C NMR (CD_3OD, 75 MHz): δ 210.1, 129.5, 129.4, 129.2, 129.0, 128.8, 128.3, 67.8, 66.6, 64.2, 34.8, 30.7, 26.1. HRMS (ESI) *m/z* calcd for $C_{18}H_{20}N_3O_4$ $(M+H)^+$: 342.1454, found: 342.1443.

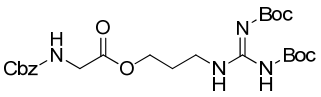
3-[_N',N''_-di(Boc)guanidino]-propanol (16)



N,N'-Di(Boc)-*S*-methylisothiourea (2.90 g, 10.0 mmol, 1 equiv) and DMAP (1.83 mg, 15.0 mmol, 1.5 equiv) were dissolved in CH_2Cl_2 (20 mL) and this mixture was cooled to 0 °C before 3-aminopropanol (765 µl, 10.0 mmol, 1 equiv) was added. After stirring for 10 min, the temperature was raised to room temperature and the mixture was stirred for 48 h. The reaction mixture was poured in citric acid (5%, 10 mL) and extracted with EtOAc (3 × 50 ml). The organic layers were combined and washed with saturated $NaHCO_3$ (10 mL) and brine (10 mL), dried over Na_2SO_4 and then evaporated to dryness. The remaining solid was purified by column chromatography (MeOH in CH_2Cl_2, 1→3%) to afford **16** (3.00 g, 95%) as an off-white solid. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 11.46 (s, NH), 8.47 (s, NH), 4.77 (s, OH), 3.70-

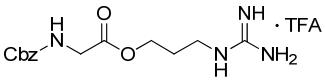
3.50 (m, 4H), 1.80-1.62 (m, 2H), 1.51 (s, 9H), 1.48 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 162.7, 156.5, 152.9, 83.1, 79.2, 57.7, 36.8, 32.5, 28.0, 27.8. HRMS (ESI) *m/z* calcd for $C_{14}H_{27}N_3NaO_5$ $(M+Na)^+$: 340.1848, found: 340.1839. Spectral data were in accordance with those reported in literature.[12]

N^{α}-Cbz-Glycine-3-[*N′,N″*-di(Boc)guanidino]-propyl ester (17)



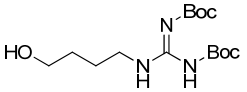
General procedure A1 using Z-Gly-OH (586 mg, 2.80 mmol), **16** (635 mg, 2.00 mmol), DMAP (49 mg, 0.20 mmol) and DCC (578 mg, 2.80 mmol). The product was obtained as a solid after purification by column chromatography (MeOH in CH_2Cl_2, 1→3%) (716 mg, 70%). R_f 0.37 (EtOAc/heptane 1:1). Mp 37 °C. IR (film) 3326, 3278, 2976, 2928, 1719, 1638, 1615, 1326, 1157, 1134, 1052 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 11.51 (s, NH), 8.44 (t, *J* = 5.2 Hz, NH), 7.39-7.27 (m, 5H), 5.44 (t, *J* = 5.5 Hz, NH), 5.13 (s, 2H), 4.24 (t, *J* = 5.8 Hz, 2H), 4.03 (d, *J* = 5.7 Hz, 2H), 3.54 (dt, *J* = 5.2, 6.3 Hz, 2H), 1.93 (tt, *J* = 6.2, 6.2 Hz, 2H), 1.50 (s, 9H), 1.49 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 170.0, 163.4, 156.3, 156.1, 153.3, 136.3, 128.4, 128.1, 128.0, 83.2, 79.3, 67.0, 63.6, 42.8, 38.2, 28.2, 28.0, 27.9. HRMS (ESI) *m/z* calcd for $C_{24}H_{37}N_4O_8$ $(M+H)^+$: 509.2611, found: 509.2598.

N^{α}-Cbz-Glycine 3-guanidinopropyl ester (Z-Gly-O3G)



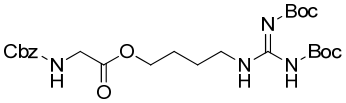
General procedure B2 using **17** (200 mg, 0.394 mmol). The product was obtained as a sticky oil (143 mg, quant). R_f 0.16 (10% MeOH in CH_2Cl_2). IR (film) 3345, 3188, 2954, 1667, 1530, 1199, 1135, 1055 cm^{-1}. ^{1}H NMR (CD_3CN, 300 MHz): δ 7.53 (br s, NH), 7.42-7.29 (m, 5H), 6.75 (br s, 3NH), 6.11 (br s, NH), 5.09 (s, 2H), 4.16 (t, *J* = 6.0 Hz, 2H), 3.84 (d, *J* = 6.2 Hz, 2H), 3.16 (dt, *J* = 6.1, 6.5 Hz, 2H), 1.86 (tt, *J* = 6.3, 6.3 Hz, 2H). ^{13}C NMR (CD_3CN, 75 MHz): δ 171.2, 158.5, 157.9, 138.0, 129.4, 128.9, 128.7, 67.3, 63.1, 43.4, 39.2, 28.5. HRMS (ESI) *m/z* calcd for $C_{14}H_{20}N_4NaO_4$ $(M+Na)^+$: 331.1382, found: 331.1378.

4-[*N′,N″*-di(Boc)guanidino]-butanol (19)



N,N′-Di(Boc)-*S*-methylisothiourea (1.06 g, 3.64 mmol, 1 equiv) and DMAP (667 mg, 5.47 mmol, 1.5 equiv) were dissolved in CH_2Cl_2 (20 mL) and this mixture was cooled to 0 °C before 4-aminobutanol (0.50 mL, 3.64 mmol, 1 equiv) was added. After stirring for 10 min, the temperature was raised to room temperature and the mixture was stirred for 48 h. The reaction mixture was poured in citric acid (5%, 10 mL) and extracted with EtOAc (3 × 50 ml). The organic layers were combined and washed with saturated $NaHCO_3$ (10 mL) and brine (10 mL), dried over Na_2SO_4 and then evaporated to dryness. The remaining solid was purified by column chromatography (EtOAc/heptane 1:1) to afford **19** (943 mg, 78%) as an off-white solid. R_f 0.25 (EtOAc/heptane 1:1). Mp 119.2 °C. IR (film) 3330, 3278, 2976, 2933, 1718, 1636, 1614, 1574, 1328, 1155, 1131, 1051, 1026 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 11.48 (s, NH), 8.37 (s, NH), 3.69 (t, *J* = 5.6 Hz, 2H), 3.44 (dt, *J* = 6.0, 6.0 Hz, 2H), 2.00 (s, OH), 1.73-1.56 (m, 4H), 1.49 (s, 18H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 163.5, 156.2, 153.3, 83.1, 79.3, 62.2, 40.3, 29.5, 28.3, 28.1, 25.6. HRMS (ESI) *m/z* calcd for $C_{15}H_{30}N_3O_5$ $(M+H)^+$: 332.2186, found: 332.2183. Spectral data were in accordance with those reported in literature.[12]

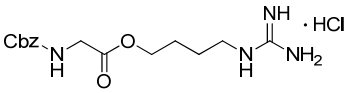
N^{α}-Cbz-Glycine-4-[*N′,N″*-di(Boc)guanidino]-butyl ester (20)



General procedure A2 using Z-Gly-OH (147 mg, 0.703 mmol), **19** (166 mg, 0.501 mmol), DMAP (12 mg, 0.10 mmol) and DCC (144 mg, 0.701 mmol). The product was obtained as an oil after purification by column chromatography (EtOAc/heptane 1:2) (250 mg, 96%). R_f 0.20 (EtOAc/heptane 1:2). IR (film) 3331, 2977, 2933, 1719, 1638, 1615, 1414, 1365, 1329,

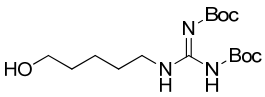
1158, 1134, 1052, 1026 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 11.49 (s, NH), 8.32 (s, NH), 7.38-7.26 (m, 5H), 5.36 (t, *J* = 5.5 Hz, NH), 5.13 (s, 2H), 4.18 (t, *J* = 6.1 Hz, 2H), 3.99 (d, *J* = 5.6 Hz, 2H), 3.44 (dt, *J* = 6.6, 6.6 Hz, 2H), 1.76-1.58 (m, 4H), 1.50 (s, 9H), 1.49 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 169.9, 163.6, 156.2, 153.3, 136.2, 128.5, 128.2, 128.1, 127.9, 83.2, 79.3, 67.0, 65.0, 42.8, 40.3, 28.3, 28.1, 25.8, 25.7. HRMS (ESI) *m/z* calcd for $C_{25}H_{38}Na_1N_4O_8$ $(M+Na)^+$: 545.2587, found: 545.2631.

N^{α}-Cbz-Glycine 4-guanidinobutyl ester (Z-Gly-O4G)



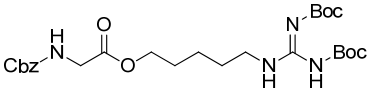
General procedure B1 using **20** (185 mg, 0.354 mmol). The product was obtained as a sticky oil (127 mg, quant).
IR (film) 3317, 3161, 2946, 1704, 1664, 1528, 1453, 1396, 1362, 1279, 1261, 1196, 1053 cm^{-1}. ^{1}H NMR (CD_3OD, 400 MHz): δ 7.38-7.27 (m, 5H), 5.11 (s, 2H), 4.19 (t, *J* = 6.1 Hz, 2H), 3.86 (s, 2H), 3.27-3.15 (m, 2H), 1.77-1.60 (m, 4H). ^{13}C NMR (CD_3OD, 75 MHz): δ 172.0, 165.6, 158.7, 138.2, 129.5, 129.1, 128.8, 67.8, 65.6, 43.5, 42.1, 26.9, 26.5. HRMS (ESI) *m/z* calcd for $C_{15}H_{22}NaN_4O_4$ $(M+Na)^+$: 345.1539, found: 345.1540.

5-[*N′,N″*-di(Boc)guanidino]-pentanol (22)



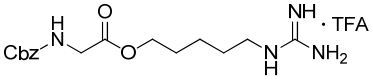
N,N′-Di-Boc-*S*-methylisothiourea (1.09 g, 3.78 mmol, 1 equiv) and DMAP (693 mg, 5.66 mmol, 1.5 equiv) were dissolved in CH_2Cl_2 (20 mL) and this mixture was cooled to 0 °C before 5-aminopentanol (404 mg, 3.92 mmol, 1 equiv) was added. After stirring for 10 min, the temperature was raised to room temperature and the mixture was stirred for 48 h. The reaction mixture was poured in citric acid (5%, 10 mL) and extracted with EtOAc (3 × 50 ml). The organic layers were combined and washed with saturated $NaHCO_3$ (10 mL) and brine (10 mL), dried over Na_2SO_4 and then evaporated to dryness. The remaining solid was purified by column chromatography (EtOAc/heptane 1:1) to afford **22** (1.3 g, 99%) as an off-white solid. R_f 0.25 (EtOAc/heptane 1:1). Mp 93.7 °C. IR (film) 3326, 3283, 2976, 2933, 1720, 1640, 1616, 1575, 1414, 1366, 1157, 1134, 1053 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 Mhz): δ 11.50 (s, NH), 8.31 (s, NH), 3.65 (t, *J* = 6.5 Hz, 2H), 3.43 (dt, *J* = 5.3, 7.0 Hz, 2H), 1.67-1.56 (m, 4H), 1.50 (s, 9H), 1.49 (s, 9H), 1.48-1.38 (m, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 163.6, 153.3, 83.1, 79.2, 62.7, 40.7, 32.3, 28.8, 28.3, 28.0, 23.0. HRMS (ESI) *m/z* calcd for $C_{16}H_{31}Na_1N_3O_5$ $(M+Na)^+$: 368.2161, found: 368.2172. Spectral data were in accordance with those reported in literature.[13]

N^{α}-Cbz-Glycine 5-[*N′,N″*-di(Boc)guanidino]pentyl ester (23)



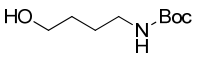
General procedure A2 using Z-Gly-OH (187 mg, 0.895 mmol), **22** (220 mg, 0.638 mmol), DMAP (15 mg, 0.123 mmol) and DCC (184 mg, 0.895 mmol). The product was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:2) (344 mg, quant). R_f 0.30 (EtOAc/heptane 1:2). IR (film) 3331, 3278, 2980, 2934, 2863, 1718, 1637, 1615, 1413, 1364, 1330, 1155, 1133, 1053, 1027 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 11.50 (s, NH), 8.32 (t, *J* = 5.4 Hz, NH), 7.40-7.28 (m, 5H), 5.27 (t, *J* = 5.5 Hz, NH), 5.13 (s, 2H), 4.16 (t, *J* = 6.5 Hz, 2H), 3.98 (d, *J* = 5.6 Hz, 2H), 3.42 (dt, *J* = 5.4, 7.1 Hz, 2H), 1.73-1.53 (m, 4H), 1.50 (s, 9H), 1.49 (s, 9H), 1.46-1.35 (m, 2H). ^{13}C NMR (CD_3OD, 75 MHz): δ 172.0, 164.5, 159.1, 157.6, 154.3, 138.2, 129.5, 129.2, 129.0, 128.9, 128.7, 84.5, 80.5, 67.8, 66.1, 43.5, 41.7, 29.7, 29.3, 28.6, 28.5, 28.3, 24.2. HRMS (ESI) *m/z* calcd for $C_{26}H_{41}N_4O_8$ $(M+H)^+$: 537.2924, found: 537.2933.

N^{α}-Cbz-Glycine 5-guanidinopentyl ester (Z-Gly-O5G)



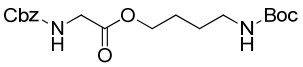
General procedure B2, using **23** (195 mg, 0.364 mmol). The product was obtained as an oily TFA salt (192 mg, 96%). IR (film) 3331, 2932, 2850, 1703, 1666, 1527, 1454, 1406, 1349, 1277, 1260, 1204, 1054 cm^{-1}. ^1H NMR (CD_3OD, 400 MHz): δ 7.38-7.27 (m, 5H), 5.11 (s, 2H), 4.17 (t, *J* = 6.4 Hz, 2H), 3.85 (s, 2H), 3.16 (t, *J* = 7.0 Hz, 2H), 1.75-1.53 (m, 4H), 1.49-1.38 (m, 2H). ^{13}C NMR (CD_3OD, 75 MHz): δ 170.2, 162.8, 161.7, 136.5, 129.5, 129.1, 128.8, 67.8, 65.9, 43.4, 42.4, 29.4, 29.3, 24.0. HRMS (ESI) *m/z* calcd for $C_{16}H_{25}N_4O_4$ $(M+H)^+$: 337.1876, found: 337.1891.

4-[*N*-(Boc)amino]-butanol (25)



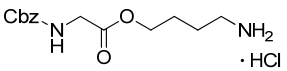
Preparation according to literature procedure.[14] 4-Aminobutanol (510 mg, 5.62 mmol, 1 equiv) and sulfamic acid (28 mg, 5 mol%) were dissolved in water (5 mL) and to this mixture di-*t*butyl dicarbonate (1.30 g, 1.1 equiv) was added in portions. After stirring for 1 h, MeOH was added and the reaction mixture was evaporated to dryness to afford **25** (1.14 g, quant) as an oily substance. Mp 25 °C. IR (film) 3353, 3278, 2980, 2934, 2868, 1682, 1528, 1424, 1365, 1273, 1250, 1167, 1041 cm^{-1}. ^1H NMR (CD_3OD, 400 MHz): δ 3.55 (t, *J* = 6.2 Hz, 2H), 3.05 (t, *J* = 6.6 Hz, 2H), 1.58-1.49 (m, 4H), 1.43 (s, 9H). ^{13}C NMR (CD_3OD, 75 MHz): δ 158.7, 79.8, 62.7, 41.2, 30.9, 28.8, 27.5. HRMS (ESI) *m/z* calcd for $C_9H_{20}N_1O_3$ $(M+H)^+$: 190.1443, found: 190.1438. Spectral data were in accordance with those reported in literature.[12]

N^{α}-Cbz-Glycine 4-[*N*-(Boc)amino]butyl ester (26)



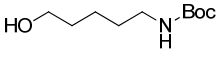
General procedure A2 using Z-Gly-OH (301 mg, 1.44 mmol), **25** (194 mg, 1.03 mmol), DMAP (26 mg, 0.210 mmol) and DCC (298 mg, 1.45 mmol). The product was obtained as a white solid after purification by column chromatography (1% MeOH in $CHCl_2$) (328 mg, 83%). R_f 0.82 (MeOH/$CHCl_2$, 9:1). Mp 83.9 °C. IR (film) 3354, 2967, 2933, 1694, 1521, 1454, 1365, 1277, 1251, 1169, 1134, 1054, 1002 cm^{-1}. ^1H NMR (CD_3OD, 400 MHz): δ 7.37-7.29 (m, 5H), 5.31 (br s, NH), 5.13 (s, 2H), 4.16 (t, *J* = 6.4 Hz, 2H), 3.97 (d, *J* = 5.6 Hz, 2H), 3.19-3.08 (m, 2H), 1.72-1.48 (m, 4H), 1.43 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 169.9, 156.2, 155.9, 136.2, 128.5, 128.2, 128.1, 79.2, 67.1, 65.1, 42.8, 40.1, 28.4, 26.6, 25.8. HRMS (ESI) *m/z* calcd for $C_{19}H_{29}Na_1N_2O_6$ $(M+Na)^+$: 403.1845, found: 403.1833.

N^{α}-Cbz-Glycine 4-aminobutyl ester (Z-Gly-O4A)



General procedure B1 using **26** (307 mg, 0.808 mmol) for the deprotection and 61 mg (0.155 mmol) for lyophilisation to give an off-white HCl salt (47 mg, 97%). Mp range 50.4 °C. IR (film) 3061, 2949, 1678, 1529, 1277, 1201, 1135, 1055, 1005 cm^{-1}. ^1H NMR (CD_3OD, 400 MHz): δ 7.39-7.27 (m, 5H), 5.11 (s, 2H), 4.19 (t, *J* = 5.7 Hz, 2H), 3.86 (s, 2H), 2.94 (t, *J* = 7.0 Hz, 2H), 1.78-1.68 (m, 4H). ^{13}C NMR (CD_3OD, 75 MHz): δ 171.9, 159.2, 138.2, 129.5, 129.1, 128.8, 67.8, 65.3, 43.4, 40.4, 26.7, 25.2. HRMS (ESI) *m/z* calcd for $C_{14}H_{21}N_2O_4$ $(M+H)^+$: 281.1501, found: 281.1486.

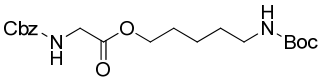
5-[*N*-(Boc)amino]-pentanol (28)



Preparation according to literature procedure.[14] 5-Aminopentanol (518 mg, 5.03 mmol, 1 equiv) and sulfamic acid (24 mg, 5 mol%) were dissolved in water (5 mL) and to this mixture di-*t*butyl-dicarbonate (1.20 mg, 1.1 equiv) was added in portions. After stirring for 1 hour, MeOH was added and the reaction mixture was evaporated to dryness to afford **28** (1.02 g, 99%) as an oil. IR (film) 3334, 2971, 2931, 2855, 1686, 1528, 1450, 1365, 1276, 1249, 1168, 1047, 1013 cm^{-1}. ^1H NMR (CD_3OD, 400 MHz): δ 3.55 (t, *J* = 6.7 Hz, 2H), 3.03 (t, *J* = 7.0 Hz, 2H), 1.62-1.32 (m, 15H). ^{13}C NMR

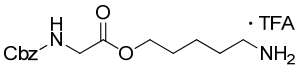
($CDCl_3$, 75 MHz): δ 156.1, 79.1, 62.6, 40.4, 32.2, 29.8, 28.4, 22.9. HRMS (ESI) *m/z* calcd for $C_{10}H_{21}Na_1N_1O_3$ $(M+Na)^+$: 226.1419, found: 226.1416. Spectral data were in accordance with those reported in literature.[15]

N^{α}-Cbz-Glycine 5-[*N*-(Boc)amino]pentyl ester (29)



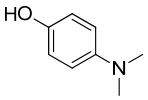
General procedure A2 using Z-Gly-OH (295 mg, 1.41 mmol), **28** (204 mg, 1.00 mmol), DMAP (25 mg, 0.205 mmol) and DCC (290 mg, 1.41 mmol). The product was obtained as a white solid after purification by column chromatography (1% MeOH in $CHCl_2$ (251 mg, 64%). R_f 0.80 (MeOH/$CHCl_2$, 9:1). Mp 51.3 °C. IR (film) 3354, 2976, 2934, 2863, 1697, 1520, 1456, 1365, 1274, 1250, 1171, 1053, 998 cm^{-1}. ^1H NMR ($CDCl_3$, 400 MHz): δ 7.27-7.38 (m, 5H), 5.36, (br s, NH), 5.12 (s, 2H), 4.14 (t, *J* = 6.6 Hz, 2H), 3.96 (d, *J* = 5.6 Hz, 2H), 3.15-3.05 (m, 2H), 1.72-1.56 (m, 2H), 1.54-1.40 (m, 11H), 1.40-1.30 (m, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 170.0, 156.2, 156.0, 136.2, 128.5, 128.2, 128.1, 79.1, 67.1, 65.3, 42.8, 40.3, 29.6, 28.4, 28.1, 23.0. HRMS (ESI) *m/z* calcd for $C_{20}H_{31}Na_1N_2O_6$ $(M+H)^+$: 417.2002, found: 417.1994.

N^{α}-Cbz-Glycine 5-aminopentyl ester (Z-Gly-O5A)



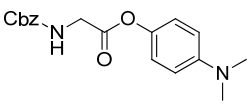
General procedure B2 using **29** (204 mg, 0.518 mmol). The product was obtained as an oily TFA salt (203 mg, quant). IR (film) 3071, 2940, 1677, 1529, 1281, 1201, 1135, 1055, 1001 cm^{-1}. ^1H NMR (CD_3OD, 400 MHz): δ 7.39-7.26 (m, 5H), 5.11 (s, 2H), 4.17 (t, *J* = 6.1 Hz, 2H), 3.86 (s, 2H), 2.91 (t, *J* = 7.6 Hz, 2H), 1.75-1.62 (m, 4H), 1.50-1.41 (m, 2H). ^{13}C NMR (CD_3OD, 75 MHz): δ 172.0, 159.2, 138.2, 129.5, 129.5, 128.8, 67.8, 65.7, 43.4, 40.6, 29.1, 28.1, 23.8. HRMS (ESI) *m/z* calcd for $C_{15}H_{23}N_2O_4$ $(M+H)^+$: 295.1659, found: 295.1650.

***p*-(Dimethylamino)phenol (31)**



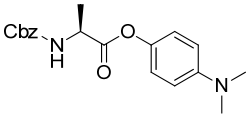
Preparation according to adapted literature procedure.[16] *p*-Aminophenol (546 mg, 5.00 mmol, 1 equiv) was dissolved in glacial acetic acid (15 mL) and heated to 40 °C under N_2. To this stirred solution paraformaldehyde (1.501 g, 50.00 mmol, 10 equiv) was added. The solution was cooled to ROOM TEMPERATURE and sodium cyanoborohydride (943 mg, 15.00 mmol, 3 equiv) was added over the course of 60 minutes. The solution was stirred overnight at 40 °C under N_2. The reaction was quenched using sodium hydroxide (20 g in 50 mL water) until pH 14 and brought back to pH 7 using 1 M HCl. After extraction with CH_2Cl_2 (8 × 30 mL), **31** (0.564 g, 82%) was obtained as an off-white solid after purification by column chromatography (MeOH in CH_2Cl_2, 1→3.5%). R_f 0.51 (10% MeOH in CH_2Cl_2). Mp 74.5-75.9 °C. IR (film) 2946, 2881, 2796, 1512, 1443, 1367, 1236, 1178, 1145, 819, 694 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 6.76 (d, *J* = 8.8 Hz, 2H), 6.70 (d, *J* = 9.3 Hz, 2H), 4.60 (s, OH), 2.86 (s, 6H).^{13}C NMR ($CDCl_3$,75 MHz): δ 148.9, 145.2, 116.2, 1161, 42.5. HRMS (ESI) *m/z* calcd for $C_8H_{12}NO$ $(M+H)^+$: 138.0919, found: 138.0941. Spectral data were in accordance with those reported in literature.[16-17]

N^{α}-Cbz-Glycine *p*-(dimethylamino)phenyl ester (Z-Gly-ODmap)



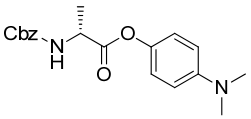
General procedure A1 using Z-Gly-OH (200 mg, 0.96 mmol), **31** (95.0 mg, 0.69 mmol), DMAP (170 mg, 1.39 mmol, 2 equiv) and DCC (173 mg, 0.84 mmol, 1.2 equiv). The product was obtained as a white solid after purification by column chromatography (2% MeOH in CH_2Cl_2) (34 mg, 15%). IR (film) 3300, 1765, 1710, 1510, 1205, 697 cm^{-1}. ^1H NMR (300 MHz, $CDCl_3$): δ 7.35-7.25 (m, 5H), 6.96 (d, *J* = 8.4 Hz, 2H), 6.68 (d, *J* = 8.6 Hz, 2H), 5.59 (br s, NH), 5.14 (s, 2H), 4.19 (d, *J* = 5.1 Hz, 2H), 2.92 (s, 6H). ^{13}C NMR (75 MHz, CD_3OD): δ 169.2, 156.2, 148.6, 140.9, 136.1, 128.3, 128.0, 127.9, 121.4, 112.8, 66.9, 42.7, 40.6. HRMS (ESI) *m/z* calcd for $C_{18}H_{21}N_2O_4$ $(M+H)^+$ 329.1501, found 329.1511.

N^{α}-Cbz-L-Alanine *p*-(dimethylamino)phenyl ester (Z-L-Ala-ODmap)



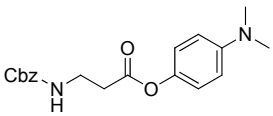
General procedure A1 using Z-L-Ala-OH (313 mg, 1.40 mmol), **31** (137 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). **Z-L-Ala-ODmap** (100 mg, 29%) was obtained as fluffy powder after purification by column chromatography (EtOAc/heptane 1:9→1:2). R_f 0.42 (EtOAc/heptane 1:2). Mp 85.8-86.2 °C. $[\alpha]_D^{20}$ +36.9 (*c* 1.00, CH_2Cl_2). IR (film) 3330, 3031, 2928, 2803, 2358, 1759, 1721, 1608, 1511, 1453, 1341, 1202, 1066, 965, 877, 810, 740, 698, 579, 530 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.39-7.28 (m, 5H), 6.95 (d, *J* = 8.9 Hz, 2 H), 6.69 (d, *J* = 9.1 Hz, 2H), 5.37 (d, *J* = 5.9 Hz, NH), 5.14 (s, 2H), 4.67-4.57 (m, 1H), 2.93 (s, 6H), 1.57 (d, *J* = 11.0 Hz, 3H). ^{13}C NMR ($CDCl_3$,75 MHz): δ 172.40, 155.74, 149.02, 141.31, 136.38, 128.68, 128.31, 128.24, 121.60, 113.13, 67.12, 49.94, 41.02, 18.94. HRMS (ESI) *m/z* calcd for $C_{19}H_{22}N_2NaO_4$ $(M+Na)^+$: 365.1477, found: 365.1481.

N^{α}-Cbz-D-Alanine *p*-(dimethylamino)phenyl ester (Z-D-Ala-ODmap)



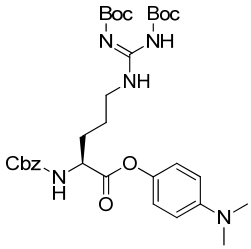
Z-D-Ala-OH (167 mg, 0.750 mmol, 1 equiv), **31** (137 mg, 1.00 mmol, 1.33 equiv) and DMAP (23 mg, 0.19 mmol, 0.25 equiv) were dissolved in dry THF and the resulting mixture was stirred under N_2 at room temperature. After the addition of Et_3N (4.60 mL, 33 mmol, 44 equiv), the mixture was cooled to 0 °C, whereupon a commercial solution of 50% propylphosphonic anhydride[4] in DMF (0.782 mL, 1.313 mmol, 1.75 equiv) was slowly added. The mixture was stirred for 30 min at 0 °C, then for 12 h at room temperature and subsequently diluted with CH_2Cl_2 (100 mL). The organic layer was washed with aq. HCl (1M, 4 × 50 mL), saturated $NaHCO_3$ (1 × 50 mL), concentrated and the residue was purified by column chromatography (EtOAc/heptane 1:9→1:2). **Z-D-Ala-ODmap** was obtained as a fluffy white powder (279 mg, 81%). R_f 0.43 (EtOAc/heptane 1:2). Mp 89.0-89.3 °C. $[\alpha]_D^{20}$ –28.6 (*c* 0.98, CH_2Cl_2). IR (film) 3363, 2942, 2885, 2794, 1764, 1703, 1608, 1515, 1453, 1344, 1292, 1204, 1116, 1064, 1014, 945, 876, 807, 747, 699, 613 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.41-7.29 (m, 5H), 6.95 (d, *J* = 8.7 Hz, 2 H), 6.69 (d, *J* = 9.0 Hz, 2H), 5.37 (d, *J* = 8.2 Hz, NH), 5.14 (s, 2H), 4.66-4.55 (m, 1H), 2.93 (s, 6H), 1.57 (d, *J* = 7.2 Hz, 3H).^{13}C NMR ($CDCl_3$,75 MHz): δ 172.4, 155.7, 149.0, 141.3, 136.4, 128.7, 128.3, 128.2, 121.6, 113.1, 67.1, 49.9, 41.0, 19.0. HRMS (ESI) *m/z* calcd for $C_{19}H_{22}N_2NaO_4$ $(M+Na)^+$: 365.1477, found: 365.1477.

N^{α}-Cbz-β-Alanine *p*-(dimethylamino)phenyl ester (Z-β-Ala-ODmap)



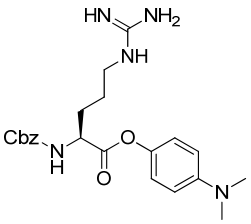
General procedure A1 using Z-β-Ala-OH (313 mg, 1.40 mmol), **31** (137 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). **Z-β-Ala-ODmap** was obtained as an off-white solid after purification by column chromatography (MeOH in CH_2Cl_2, 1→4%) (314 mg, 92%). R_f 0.34 (EtOAc/heptane 1:2). Mp 71.8-73.9 °C. IR (film) 2931, 1750, 1718, 1515, 1208, 1154, 1006, 699 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz):δ 7.29-7.37 (m, 5H), 6.93 (d, *J* = 9.2 Hz, 2H), 6.69 (d, *J* = 9.2 Hz, 2H), 5.29 (d, *J* = 7.1 Hz, NH) 5.11 (s, 2H), 3.56 (dt, *J* = 6.1, 12.1 Hz, 2H), 2.93 (s, 6H), 2.78 (t, *J* = 5.9 Hz, 2H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 171.7, 156.4, 148.9, 141.3, 136.6, 128.7, 128.2, 128.2, 121.8, 113.2, 66.9, 41.0, 36.7, 34.7. HRMS (ESI) *m/z* calcd for $C_{19}H_{22}N_2NaO_4$ $(M+Na)^+$: 365.1477, found: 365.1481.

N^{α}-Cbz-L-Arginine(diBoc) *p*-(dimethylamino)phenyl ester (32R)



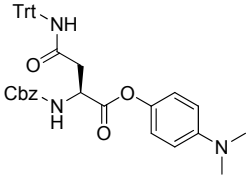
General procedure A2 using Z-L-Arg(diBoc)-OH (254 mg, 0.50 mmol, 1 equiv), **31** (69 mg, 0.50 mmol), DMAP (12 mg, 0.10 mmol) and DCC (144 mg, 0.70 mmol). **32R** was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:3→1:2) (270 mg, 86%). R_f 0.30 (EtOAc/heptane 1:2). Mp 47 °C. $[\alpha]_D^{20}$ +6.1 (*c* 1.47, CH_2Cl_2). IR (film) 3334, 3283, 2967, 2924, 1753, 1716, 1636, 1613, 1513, 1329, 1131, 1050 cm^{-1}. ^1H NMR (300 MHz, $CDCl_3$) δ 11.55 (s, NH), 8.26 (br t, *J* = 5.7 Hz, NH), 7.41-7.27 (m, 5H), 6.93-6.83 (m, 2H), 6.76-6.67 (m, 2H), 6.25 (d, *J* = 7.1 Hz, NH), 5.11 (s, 2H), 4.44-4.35 (m, 1H), 3.48-3.29 (m, 2H), 2.90 (s, 6H), 1.92-1.58 (m, 4H), 1.48 (s, 9H), 1.42 (s, 9H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 172.7, 164.6, 157.3, 153.9, 150.0, 142.2, 129.4, 128.9, 128.7, 122.6, 113.8, 83.9, 79.4, 67.1, 55.1, 41.0, 40.7, 29.3, 28.4, 28.4, 28.1, 26.3. HRMS (ESI) *m/z* calcd for $C_{32}H_{46}N_5O_8$ $(M+H)^+$: 628.3346, found: 628.3304.

N^{α}-Cbz-L-Arginine *p*-(dimethylamino)phenyl ester (Z-L-Arg-ODmap)



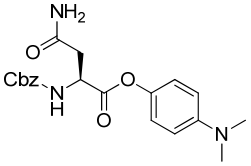
To a stirred solution of **32R** (100 mg, 0.16 mmol, 1 equiv) in CH_2Cl_2 (2 mL), TFA (0.5 mL) and triethylsilane (80 µL) were added and the resulting mixture was stirred for 1 hour. The solvent was removed under reduced pressure, redissolved in water/dioxane (1:2 v/v) and washed with heptane. The waterlayer was lyophilised in the presence of 2M HCl to obtain the product as an off-white HCl salt (70 mg, quant). Mp 42.5 °C. $[\alpha]_D^{20}$ –19.6 (*c* 1.95, MeOH). IR (film) 3330, 3261, 3170, 2954, 1765, 1668, 1509, 1203, 1175, 1130 cm^{-1}. ^1H NMR (300 MHz, CD_3CN) δ 7.74 (br s, NH), 7.65-7.57 (m, 2H), 7.41-7.29 (m, 5H), 7.24-7.16 (m, 2H), 6.66-6.50 (br m, 4NH), 5.11 (s, 2H), 4.44-4.34 (m, 1H), 3.24-3.11 (m, 2H), 3.08 (s, 6H), 2.10-1.68 (m, 4H, excl CD_3CN). ^{13}C NMR (CD_3CN, 75 MHz) δ 171.9, 158.4, 157.4, 150.4, 143.2, 138.0, 129.4, 128.9, 128.7, 124.0, 122.0, 67.3, 55.0, 45.9, 41.7, 29.0, 25.7. HRMS (ESI) *m/z* calcd for $C_{22}H_{30}N_5O_4$ $(M+H)^+$: 428.2298, found: 428.2288.

N^{α}-Cbz-L-Asparagine(Trt) *p*-(dimethylamino)phenyl ester (32N)



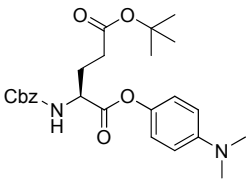
Z-L-Asn(Trt)-OH (678 mg, 1.33 mmol, 1.33 equiv), **31** (137 mg, 1.00 mmol, 1 equiv) and DMAP (41 mg, 0.33 mmol, 0.3 equiv) were dissolved in dry THF and the resulting mixture was stirred under N_2 at room temperature. After the addition of Et_3N (8.18 mL, 58.7 mmol, 44 equiv), the mixture was cooled to 0 °C, whereupon a commercial solution of 50% propylphosphonic anhydride[4] in DMF (1.389 mL, 2.33 mmol, 1.75 equiv) was slowly added. The mixture was stirred for 30 min at 0 °C, then for 12 h at room temperature and subsequently diluted with CH_2Cl_2 (100 mL). The organic layer was washed with aq. HCl (1M, 4 × 20 mL), saturated $NaHCO_3$ (1 × 25 mL), concentrated and the residue was purified by column chromatography (EtOAc/heptane 1:5→1:2). **32N** was obtained as a white powder (333 mg, 40%). R_f 0.31 (EtOAc/heptane 1:2). Mp 179.3-188.2 °C. $[\alpha]_D^{20}$ +15.8 (*c* 0.99, CH_2Cl_2). IR (film) 3316, 3029, 1763, 1718, 1511, 1447, 1336, 1261, 1204, 1129, 1045, 902, 804, 750, 700, 623, 614, 605 cm^{-1}. ^1H NMR (300 MHz, $CDCl_3$) δ 7.40-7.11 (m, 20H), 6.78 (d, *J* = 9.1 Hz, 2H), 6.64 (d, *J* = 9.2 Hz, 2H), 6.11 (d, *J* = 7.4 Hz, NH), 5.13 (s, 2H), 4.80-4.73 (m, 1H), 3.22 (dd, *J* = 15.7, 4.7 Hz, 1H), 2.95 (dd, *J* = 15.7, 4.7 Hz, 1H), 2.93 (s, 6H). ^{13}C NMR ($CDCl_3$,75 MHz) δ 170.5, 169.2, 156.4, 150.0, 144.4, 141.7, 136.5, 128.8, 128.6, 128.2, 128.0, 127.3, 121.9, 113.2, 71.1, 67.1, 51.1, 41.1, 39.0. HRMS (ESI) *m/z* calcd for $C_{39}H_{38}N_3O_5$ $(M+H)^+$: 628.2812, found: 628.2813.

N^{α}-Cbz-L-Asparagine *p*-(dimethylamino)phenyl ester (Z-L-Asn-ODmap)



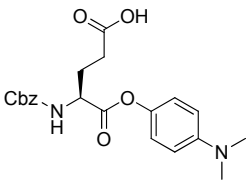
To a stirred solution of **32N** (150 mg, 0.24 mmol, 1 equiv) in CH_2Cl_2 (2 mL), TFA (552 µL, 7.17 mmol, 30 equiv) and triethylsilane (43 µL, 0.27 mmol, 1.13 equiv) were added and the resulting mixture was stirred overnight. After precipitation with Et_2O, **Z-L-Asn-ODmap** was obtained as a white powder (46 mg, 50%). R_f 0.47 (10% MeOH in CH_2Cl_2). Mp 136.2-139.3 °C. $[\alpha]_D^{20}$ –3.1 (*c* 0.14, CH_2Cl_2). IR (film) 3316, 1757, 1694, 1655, 1516, 1446, 1281, 1205, 1058, 808, 698, 613, 603, 417 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.39-7.19 (m, 5H), 7.00 (d, *J* = 8.8 Hz, 2H), 6.83 (d, *J* = 9.0 Hz, 2H), 6.10 (d, *J* = 8.4 Hz, NH), 5.63 (d, *J* = 13.9 Hz, NH_2), 5.14 (s, 2H), 4.85-4.74 (m, 1H), 2.95 (s, 6H), 3.19-2.80 (m, 2H). ^{13}C NMR ($CDCl_3$,75 MHz): δ 172.1, 170.5, 156.4, 147.7, 143.3, 136.3, 129.6, 128.7, 128.4, 128.3, 128.2, 126.4, 122.1, 114.8, 67.3, 50.9, 42.0, 37.6. HRMS (ESI) *m/z* calcd for $C_{20}H_{23}N_3NaO_5$ $(M+Na)^+$: 408.1535, found: 408.1548.

N^{α}-Cbz-L-Glutamate (O^{t}Bu) *p*-(dimethylamino)phenyl ester (32E)



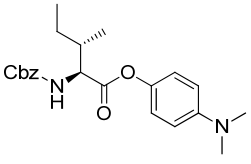
General procedure A1 using Z-L-Glu(O^{t}Bu)-OH (472 mg, 1.40 mmol), **31** (137 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). **32E** was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:9→1:2) (442 mg, 97%). R_f 0.48 (EtOAc/heptane 1:2). Mp 62.2-64.2 °C. $[\alpha]_D^{20}$ –12.7 (*c* 0.99, CH_2Cl_2). IR (film) 3338, 2976, 1724, 1609, 1511, 1453, 1203, 1151, 1057, 947, 811, 739, 698, 512 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.38-7.27 (m, 5H), 6.95 (d, *J* = 8.8 Hz, 2H), 6.68 (d, *J* = 9.1 Hz, 2H), 5.48 (d, *J* = 7.7 Hz, NH), 5.13 (s, 2H), 4.66-4.57 (m, 1H), 2.93 (s, 6H), 2.50-2.38 (m, 1H), 2.37-2.23 (m, 1H), 2.18-2.02 (m, 1H), 1.44 (s, 9H).^{13}C NMR ($CDCl_3$,75 MHz): δ 149.1, 141.2, 128.7, 128.3, 128.3, 121.7, 113.1, 67.2, 53.8, 41.0, 31.6, 28.2, 27.9. HRMS (ESI) *m/z* calcd for $C_{25}H_{32}N_2NaO_6$ $(M+Na)^+$: 479.2158, found: 479.2154.

N^{α}-Cbz-L-Glutamate *p*-(dimethylamino)phenyl ester (Z-L-Glu-ODmap)



General procedure B2 using **32E** (200 mg, 0.438 mmol). **Z-L-Glu-ODmap** was obtained as a thick dark yellow oil (175 mg, quant). R_f 0.44 (10% MeOH in CH_2Cl_2). $[\alpha]_D^{20}$ –10.1 (*c* 0.80, CH_2Cl_2). IR (film) 3308, 2962, 1766, 1705, 1671, 1508, 1455, 1398, 1174, 1174, 1130, 1058, 836, 798, 722, 700, 607, 553 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.52 (d, *J* = 8.5 Hz, 2H), 7.40-7.30 (m, 5H), 7.24 (d, *J* = 10.6 Hz, 2H), 6.19 (s, OH), 5.62 (d, *J* = 6.4 Hz, NH), 5.13 (s, 2H), 4.70-4.58 (m, 1H), 3.17 (s, *J* = 6.2 Hz, 6H), 2.56 (t, *J* = 6.6 Hz, 1H), 2.44 – 2.28 (m, 1H), 2.24 – 2.09 (m, 1H). ^{13}C NMR ($CDCl_3$,75 MHz): δ 176.3, 161.3, 156.4, 150.7, 141.1, 128.8, 128.5, 128.3, 123.7, 121.7, 67.6, 53.6, 46.5, 31.2, 29.7, 26.9. HRMS (ESI) *m/z* calcd for $C_{21}H_{24}N_2NaO_6$ $(M+Na)^+$: 423.1553, found: 423.1532.

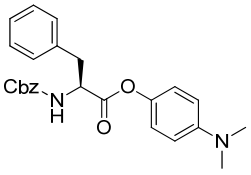
N^{α}-Cbz-L-Isoleucine *p*-(dimethylamino)phenyl ester (Z-L-Ile-ODmap)



General procedure A1 using Z-L-Ile-OH (371 mg, 1.40 mmol), **31** (137 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). **Z-L-Ile-ODmap** was obtained as an off-white solid after purification by column chromatography (EtOAc/heptane 1:9→1:2) (375 mg, 98%). R_f 0.56 (EtOAc/heptane 1:2). Mp 90.6-91.7 °C. $[\alpha]_D^{20}$ –7.3 (*c* 0.94, CH_2Cl_2). IR (film) 2963, 2928, 1720, 1513, 1206, 1163, 698 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.39-7.31 (m, 5H), 6.94 (d, *J* = 8.9 Hz, 2H), 6.69 (d, *J* = 9.1 Hz, 2H), 5.36 (d, *J* = 8.8 Hz, NH), 5.14 (s, 2H), 4.57 (dd, *J* = 4.8, 8.9 Hz, 1H), 2.93 (s, 6H), 2.04 (s,

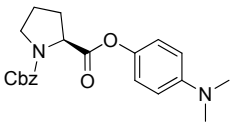
1H), 1.63-1.50 (m, 2H), 1.04 (d, J = 6.9 Hz, 3H), 0.98 (t, J = 7.3 Hz, 3H).13C NMR (CDCl3,75 MHz): δ 171.4, 156.3, 149.0, 141.3, 136.4, 128.7, 128.3, 128.3, 121.7, 113.2, 67.2, 58.5, 41.0, 38.5, 25.3, 15.7, 11.9. HRMS (ESI) *m/z* calcd for $C_{22}H_{28}N_2NaO_4$ $(M+Na)^+$: 407.1947, found: 407.1951.

N^{α}-Cbz-L-Phenylalanine *p*-(dimethylamino)phenyl ester (Z-L-Phe-ODmap)



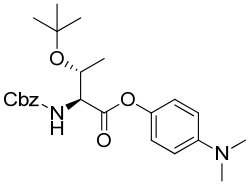
General procedure A1 using Z-L-Phe-OH (419 mg, 1.40 mmol), **31** (137 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). **Z-L-Phe-ODmap** was obtained as a white fluffy powder after purification by column chromatography (EtOAc/heptane 1:9→1:2) (376 mg, 90%). R_f 0.47 (EtOAc/heptane 1:2). Mp 79.8-80.4 °C. $[\alpha]_D^{20}$ –9.3 (*c* 1.00, CH_2Cl_2). IR (film) 1721, 1515, 1348, 1204, 1054, 746, 699, 611, cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.37 – 7.19 (m, 10H), 6.84 (s, 2H), 6.67 (d, *J* = 9.0 Hz, 2H), 5.31 (d, *J* = 8.1 Hz, NH), 5.12 (s, 2H), 4.88 (dt, *J* = 13.4, 6.0 Hz, 1H), 3.26 (d, *J* = 6.2 Hz, 2H), 2.92 (s, 6H).^{13}C NMR ($CDCl_3$,75 MHz): δ 170.7, 155.6, 148.8, 141.1, 136.2, 135.5, 129.4, 128.5, 128.2, 128.1, 127.2, 121.4, 112.9, 67.0, 40.8, 38.3. HRMS (ESI) *m/z* calcd for $C_{25}H_{26}N_2NaO_4$ $(M+Na)^+$: 441.1790, found: 441.1800.

N^{α}-Cbz-L-Proline *p*-(dimethylamino)phenyl ester (Z-L-Pro-ODmap)



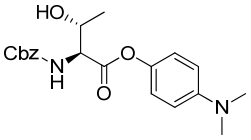
General procedure A1 using Z-L-Pro-OH (349 mg, 1.40 mmol), **31** (137 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). **Z-L-Pro-ODmap** was obtained as a white solid after purification by column chromatography (EtOAc/heptane 1:9→1:2) (325 mg, 88%). R_f 0.35 (EtOAc/heptane 1:2). MP 71.1-72.0 $[\alpha]_D^{20}$ –88.7 (*c*1.00, CH_2Cl_2). IR (film) 3321, 3274, 3058, 3028, 2353, 1692, 1640, 1527, 1493, 1450, 1290, 1238, 1048, 996, 745, 694, 625, 569 cm^{-1}.^1H NMR ($CDCl_3$, 300 MHz): δ 7.44-7.23 (m, 5H), 6.97 (d, *J* = 9.1 Hz, 2H), 6.71 (d, *J* = 8.9 Hz, 2H), 5.30-5.07 (m, 2H), 4.63-4.48 (m, 1H), 3.85-3.47 (m, 2H), 2.91 (s, 6H), 2.45-2.12 (m, 2H), 2.11 (s, 2H) .^{13}C NMR ($CDCl_3$,75 MHz): δ 172.0, 171.9, 155.0, 154.4, 148.9, 141.7, 141.4, 136.9, 136.6, 128.6, 128.1, 128.0, 121.8, 121.6, 113.2, 113.1, 67.3, 67.2, 59.5, 59.1, 47.2, 46.6, 41.0, 31.2, 30.2, 24.5, 23.7. (HRMS (ESI) *m/z* calcd for $C_{21}H_{24}N_2NaO_4$ $(M+Na)^+$: 391.1634, found: 391.1639.

N^{α}-Cbz-L-Threonine (O^tBu) *p*-(dimethylamino)phenyl ester (32T)



General procedure A1 using Z-L-Thr(*t*-Bu)-OH (433 mg, 1.40 mmol), **31** (137 mg, 1.00 mmol), DMAP (24 mg, 0.20 mmol) and DCC (289 mg, 1.40 mmol). **32T** was obtained as a white solid after purification by column chromatography using (EtOAc/heptane 1:9→1:2) (358 mg, 86%). R_f 0.45 (EtOAc/heptane 1:2). Mp 90.8-91.7 °C. $[\alpha]_D^{20}$ –17.9 (*c* 1.00, CH_2Cl_2). IR (film) 3446, 2975, 1765, 1725, 1609, 1511, 1454, 1346, 1309, 1195, 1091, 1068, 969, 739, 698, 511 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz):δ 7.42-7.27 (m, 5H), 6.97 (d, *J* = 9.1 Hz, 2H), 6.69 (d, *J* = 9.2 Hz, 2H), 5.67 (d, *J* = 9.5 Hz, NH), 5.15 (s, 2H), 4.45-4.38 (m, 1H), 2.92 (s, 6H), 1.28 (d, *J* = 13.8 Hz, 3H), 1.17 (s, 9H), 0.89-0.84 (m, 1H). ^{13}C NMR ($CDCl_3$,75 MHz): δ 170.4, 157.0, 148.9, 141.6, 136.5, 128.7, 128.3, 121.7, 113.2, 74.3, 67.5, 67.2, 60.3, 41.1, 28.8, 24.8, 21.2.HRMS (ESI) *m/z* calcd for $C_{24}H_{32}N_2NaO_5$ $(M+Na)^+$: 451.2209, found: 451.2218.

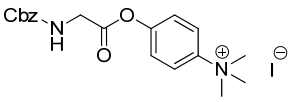
N^{α}-Cbz-L-Threonine *p*-(dimethylamino)phenyl ester (Z-L-Thr-ODmap)



General procedure B1 using **32T** (100 mg, 0.233 mmol). Instead of lyophilisation, the resulting mixture was purified by column chromatography (EtOAc/heptane, 5:1→1:5). **Z-L-Thr-ODmap** was obtained as a brown oil (87 mg, quant). $[\alpha]_D^{20}$ –24.7 (*c* 0.40, CH_2Cl_2).

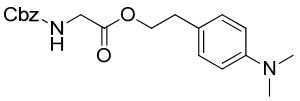
IR (film) 3375, 2925, 1761, 1677, 1609, 1513, 1455, 1202, 1131, 1066, 1006, 836, 799, 722, 698, 552 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz: δ 7.42-7.28 (m, 5H), 7.02 (dd, *J* = 21.5, 9.1 Hz, 4H), 5.70 (d, *J* = 8.9 Hz, 1H), 5.16 (s, 2H), 4.64-4.45 (m, *J* = 7.4 Hz, 3H), 2.99 (s, 6H), 1.33 (d, *J* = 6.3 Hz, 3H). ^{13}C NMR ($CDCl_3$,75 MHz): δ 170.3, 146.4, 136.3, 128.7, 128.4, 128.2, 122.4, 116.3, 68.3, 67.5, 59.5, 42.9, 29.8, 20.3, 14.3. HRMS (ESI) *m/z* calcd for $C_{21}H_{24}N_2NaO_6$ $(M+Na)^+$: 395.1583, found: 395.1583.

N^{α}-Cbz-Glycine *p*-(trimethylammonium)phenyl ester (Z-Gly-OTmap)



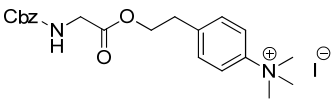
Preparation according to literature procedure.[18] To a solution of **Z-Gly-ODmap** (40 mg, 0.12 mmol, 1 equiv) in MeCN (5 mL) was added methyl iodide (0.30 mL, 4.8 mmol, 40 equiv). The mixture was refluxed for at least 2 hours. Afterwards, the solvent and methyl iodide were evaporated and the resulting mixture was purified with column chromatography (EtOAc) to afford a white powder (34 mg, 59%). IR (film) 1765, 1711, 1512, 1160, 697 cm^{-1}. ^1H NMR (300 MHz, CD_3OD): δ 7.99 (d, *J* = 9.5 Hz, 2H), 7.43 (d, *J* = 9.4 Hz, 2H), 7.33 (m, 5H), 5.14 (s, 2H), 4.16 (s, 2H), 3.69 (s, 9H). ^{13}C NMR (75 MHz, CD_3OD): δ 176.7, 171.7, 143.6, 132.1, 129.5, 129.5, 129.0, 158.7, 121.1, 120.9, 67.7, 57.8, 43.4. HRMS (ESI) *m/z* calcd for $C_{19}H_{23}N_2O_4^+$ $(M)^+$ 343.1658, found 343.1641.

N^{α}-Cbz-Glycine *p*-(dimethylamino)phenethyl ester (Z-Gly-ODmape)



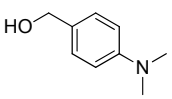
General procedure A1 using Z-Gly-OH (146 mg, 0.70 mmol), *p*-(dimethylamino)phenethyl alcohol (82 mg, 0.50 mmol), DMAP (122 mg, 1.0 mmol, 2 equiv) and DCC (144 mg, 0.7 mmol) in EtOAc (3 mL). The product was obtained by silica gel column purification (2% MeOH in CH_2Cl_2) (132 mg, 75%). IR (film) 3300, 1720, 1520, 1189, 698 cm^{-1}. ^1H NMR (300 MHz, $CDCl_3$): δ 7.34 (m, 5H) 7.07 (d, *J* = 8.6 Hz, 2H,) 6.68 (d, *J* = 8.7 Hz, 2H), 5.22 (br s, NH), 5.13 (s, 2H), 4.30 (t, *J* = 7.2 Hz, 2H) 3.97 (d, *J* = 5.5 Hz, 2H), 2.91 (s, 6H), 2.85 (t, *J* = 7.1 Hz, 2H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 169.9, 149.5, 135.8, 129.5, 128.5, 128.2, 128.1, 125.0, 122.9, 67.1, 66.4, 42.8, 40.7, 34.0. HRMS (ESI+) *m/z* calcd for $C_{20}H_{25}N_2O_4$ $(M+H)^+$ 357.1814, found 357.1710.

N^{α}-Cbz-Glycine *p*-(trimethylammonium)phenethyl ester (Z-Gly-OTmape)



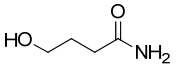
To a solution of **Z-Gly-ODmape** (77 mg, 0.22 mmol, 1 equiv) in MeCN (5 mL) was added methyl iodide (0.30 mL, 4.8 mmol, 22 equiv). The mixture was refluxed for at least 2 hours. Afterwards, the solvent and methyl iodide were evaporated and the resulting mixture was purified with column chromatography (EtOAc) to afford a light yellow powder (110 mg, quant). ^1H NMR (300 MHz, CD_3OD): δ 7.81 (d, *J* = 9.0 Hz, 2H), 7.52 (d, *J* = 9.0 Hz, 2H), 7.32 (m, 5H), 5.10 (s, 2H), 4.38 (t, *J* = 6.5 Hz, 2H,), 3.80 (s, 2H), 3.63 (s, 9H), 3.04 (t, *J* = 6.5 Hz, 2H). ^{13}C NMR (75 MHz, CD_3OD): δ 171.7, 151.7, 142.6, 132.2, 129.5, 129.5, 129.1, 128.7, 121.1, 120.9, 67.8, 66.0, 57.8, 43.4, 35.2. HRMS (ESI) *m/z* calcd for $C_{21}H_{27}N_2O_4^+$ $(M)^+$ 371.1974, found 371.1978.

***p*-(Dimethylamino)benzylalcohol (35)**



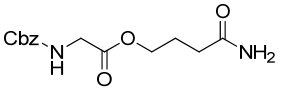
To a solution of *p*-(dimethylamino)benzaldehyde (149 mg, 1.00 mmol, 1 equiv) in MeOH (2.5 mL) was added sodium borohydride (45 mg, 1.2 mmol, 1.2 equiv,). After stirring for 10 minutes, the mixture was concentrated, dissolved in EtOAc and washed with water (3 × 5 mL) and concentrated again *in vacuo* to afford product **35** (121 mg, 80%). ^1H NMR (300 MHz, $CDCl_3$): δ 7.19 (d, *J* = 8.8 Hz, 2H), 6.69 (d, *J* = 8.8 Hz, 2H), 4.49 (s, 2H), 2.90 (s, 6H), 2.32 (br s, OH). HRMS (ESI+) *m/z* calcd for $C_9H_{13}NO$ $(M+H)^+$ 151.10, found . Spectral data were in accordance with those reported in literature.[19]

γ-Hydroxybutyramide (37)



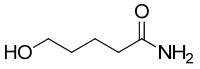
Preparation according to literature procedure.[20] To cooled (–78 °C) γ-butyrolactone (5.00 g, 58.1 mmol, 1.0 equiv) in a sealed tube was added liquid ammonia (40 mL). The tube was warmed to 40 °C and stirred for 5 days. The excess ammonia was evaporated and the crude product was recrystallised from EtOAc to afford **37** (5.81 g, 97%) as white needles that liquified upon standing. R_f 0.40 (5% MeOH in CH_2Cl_2). IR (film) 3342, 3197, 2941, 2858, 1653, 1605, 1404, 1051, 926, 580 cm^{-1}. ^1H NMR (CD_3OD, 300 MHz): δ 3.57 (t, *J* = 6.6 Hz, 2H), 2.29 (t, *J* = 7.5 Hz, 2H), 1.81 (dt, *J* = 6.6, 7.5 Hz, 2H). ^{13}C NMR (CD_3OD, 75 MHz): δ 210.1, 128.8, 64.6, 30.7. HRMS (ESI) *m/z* calcd for $C_4H_{10}NO_2$ $(M+H)^+$: 104.0712 found: 104.0716. Spectral data were in accordance with those reported in literature.[21]

***N*α-Cbz-Glycine 3-carbamoylpropyl ester (Z-Gly-O3Cam)**



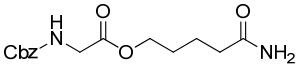
A solution of **37** (242 mg, 2.3 mmol, 1.1 equiv) in CH_2Cl_2 (4 mL) was cooled to 0 °C. Subsequently, EDC·HCl (408 mg, 2.1 mmol, 1.0 equiv), DMAP (260 mg, 2.1 mmol, 1.0 equiv) and Cbz-Gly-OH (446 mg, 2.1 mmol, 1.0 equiv) were added. The reaction mixture was allowed to slowly warm to room temperature and stirred during 6 hours. Work-up was performed by diluting with CH_2Cl_2 (10 mL), washing with 5% $NaHCO_3$ (2 × 5 mL) and brine (1 × 5 mL), drying with Na_2SO_4, filtration and evaporation to dryness. The crude product was purified by column chromatography (MeOH in CH_2Cl_2, 1→4%) to afford **Z-Gly-O3Cam** (286 mg, 46%) as a white solid. R_f 0.56 (5% MeOH in CH_2Cl_2). IR (film) 3349, 2920, 2553, 1708, 1667, 1646, 1625, 1522, 1432, 1348, 1272, 1189, 1058, 1009, 608 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.36 (m, 5H), 5.58 (br s, NH), 5.30 (br s, 2NH), 5.13 (s, 2H), 4.23 (t, *J* = 6.0 Hz, 2H), 3.96 (s, 2H), 2.28 (t, *J* = 2.7 Hz, 2H), 1.98 (dt, *J* = 6.0, 2.7 Hz, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 173.9, 136.1, 128.6, 128.3, 128.0, 67.2, 64.4, 42.9, 42.8, 31.7, 30.9, 24.3. HRMS (ESI) *m/z* calcd for $C_{14}H_{18}N_2NaO_5$ $(M+Na)^+$: 317.1113, found: 317.1106.

δ-Hydroxyvaleramide (39)



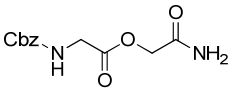
Preparation according to literature procedure.[20] To cooled (–78 °C) δ-valerolactone (1.03 g, 10.3 mmol, 1.0 equiv) in a sealed tube was added liquid ammonia (20 mL). The tube was warmed to 40 °C and stirred for 5 days. The excess ammonia was evaporated and the crude product was purified by column chromatography (MeOH in CH_2Cl_2, 3→20%) (746 mg, 62%). ^1H NMR (CD_3CN, 300 MHz): δ 6.06 (br s, NH), 5.54 (br s, NH), 3.48 (dt, *J* = 5.5, 6.2 Hz, 2H), 2.60 (t, *J* = 5.5 Hz, OH), 2.15 (t, *J* = 7.3 Hz, 2H), 1.65-1.42 (m, 4H). ^{13}C NMR (CD_3CN, 75 MHz): δ 175.9, 62.1, 35.7, 33.0, 22.5. HRMS (ESI) *m/z* calcd for $C_5H_{11}NO_2$ $(M+H)^+$: 118.0868, found: 118.0869. Spectral data were in accordance with those reported in literature.[22]

***N*α-Cbz-Glycine 4-carbamoylbutyl ester (Z-Gly-O4Cam)**



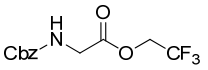
General procedure A1 using Z-Gly-OH (586 mg, 2.80 mmol) **39** (234 mg, 2.00 mmol), DMAP (49 mg, 0.4 mmol) and DCC (578 mg, 2.80 mmol). The product was obtained as a white solid after purification by column chromatography (MeOH in CH_2Cl_2, 3→8%) (578 mg, 94%). R_f 0.24 (5% MeOH in CH_2Cl_2). Mp 77 °C. IR (film) 3334, 2950, 1709, 1662, 1524, 1282, 1195, 1057 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.41-7.28 (m, 5H), 5.60 (br s, NH), 5.53 (br s, NH), 5.38 (t, *J* = 5.7 Hz, 2H), 5.13 (s, 2H), 4.18 (t, *J* = 5.9 Hz, 2H), 3.96 (d, *J* = 5.7 Hz, 2H), 2.24 (t, *J* = 6.9 Hz, 2H), 1.76-1.65 (m, 4H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 174.8, 170.0, 156.4, 136.2, 128.5, 128.2, 128.0, 67.1, 64.9, 42.8, 35.0, 27.9, 21.8. HRMS (ESI) *m/z* calcd for $C_{15}H_{21}N_2O_5$ $(M+H)^+$: 309.1451, found: 309.1444.

N^{α}-Cbz-Glycine carbamoylmethyl ester (Z-Gly-OCam)



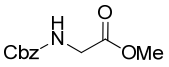
Preparation according to literature procedure.[5] Z-Gly-OH (2.50 g, 12.0 mmol, 1 equiv) was dissolved in EtOH (10 mL) and diluted with water (3 mL) and KOH (671 mg, 12.0 mmol, 1 equiv) was added to form the potassium salt. After 5 minutes the mixture was concentrated *in vacuo* and co-evaporated with toluene (3 × 20 mL). The residue was dissolved in dry DMF (6 mL) and chloroacetamide (1.34 g, 14.3 mmol, 1.2 equiv) was added. The mixture was stirred overnight at 50°C. During the reaction KCl precipitated as a white powder. After filtration, DMF was removed *in vacuo*, the product was dissolved in EtOAc and washed with a saturated $NaHCO_3$ solution (3 × 20 mL), and water (3 × 20 mL). The organic layer was dried over Na_2SO_4, the solvent was removed *in vacuo* and the product was obtained as a white powder (1.57 g, 49%). ^{1}H NMR (400 MHz, CD_3OD): δ 7.32-7.27 (m, 5H), 5.11 (s, 2H), 4.60 (s, 2H), 4.00 (s, 2H). ^{13}C NMR (75 MHz, CD_3OD) δ (ppm) 178.4, 172.5, 159.1, 138.1, 129.5, 129.0, 128.8, 67.8, 62.4, 43.3 HRMS (ESI) *m/z* calcd for $C_{12}H_{14}N_2NaO_5$ $(M+Na)^+$: 289.0800, found: 289.0810. Spectral data were in accordance with those reported in literature.[23]

N^{α}-Cbz-Glycine [2,2,2-trifluoro]ethyl ester (Z-Gly-OTfe)



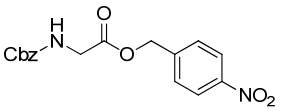
General procedure A3 using Z-Gly-OH (1.40 g, 6.7 mmol), DMAP (819 mg, 6.7 mmol), trifluoroethanol (0.48 mL, 6.7 mmol) and EDC (1.54 g, 8.0 mmol. The product was obtained as a white powder after purification by column chromatography (EtOAc/heptane 1:1) (1.25 g, 64%). ^{1}H NMR (300 MHz, $CDCl_3$): δ 7.37-7.29 (m, 5H), 5.21 (br s, NH), 5.14 (s, 2H), 4.53 (q, *J* = 8.2 Hz, 2H), 4.10 (d, *J* = 5.9 Hz, 2H). ^{13}C NMR (75 MHz, $CDCl_3$): δ 168.7, 156.2, 136.0, 128.6, 128.3, 128.1, 122.6 (q, *J* = 278 Hz, 1C), 67.3, 61.9 (q, *J* = 37 Hz, 1C), 42.2. HRMS (ESI) *m/z* calcd for $C_{12}H_{12}F_3NNaO_4$ $(M+Na)^+$: 314.0616, found: 314.0613.

N^{α}-Cbz-Glycine methyl ester (Z-Gly-OMe)



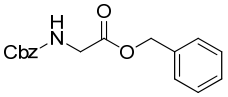
To a solution of Z-Gly-OH (2.09 g, 10 mmol, 1 equiv) in dry THF (10 mL) at 0°C, triethylamine (1.39 mL, 10 mmol, 1 equiv) was added as well as methylchloroformate (1.40 mL, 18 mmol, 1.8 equiv). The mixture was concentrated and diluted in MeOH. This mixture was refluxed for 1 hour, concentrated and purified by column chromatography (2.14 g, 96%). ^{1}H NMR (300 MHz, CD_3OD): δ 7.35-7.20 (m, 5H), 5.09 (s, 2H), 3.85 (s, 2H), 3.69 (s, 3H). ^{13}C NMR (CD_3OD, 75 MHz): δ 172.4, 159.2, 138.2, 129.5, 129.1, 128.8, 67.8, 52.6, 43.3. HRMS (ESI) *m/z* calcd for $C_{11}H_{13}NNaO_4$ $(M+Na)^+$: 246.0742, found: 246.0752. Spectral data were in accordance with those reported in literature.[24]

N^{α}-Cbz-Glycine *p*-nitrobenzyl ester (Z-Gly-ONb)



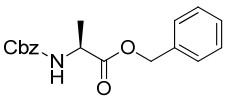
General procedure A2 using Z-Gly-OH (295 mg, 1.41 mmol), *p*-nitrobenzyl alcohol (155 mg, 1.01 mmol), DMAP (25 mg, 0.20 mmol) and DCC (300 mg, 1.46 mmol). The product was obtained as an off- white solid after purification by column chromatography (EtOAc/heptane 1:2) (290 mg, 83%). R_f 0.25 (EtOAc/heptane 1:2). Mp 109.5 °C. IR (film) 3347, 2928, 1736, 1717, 1513, 1351, 1212, 1011, 734 cm^{-1}. ^{1}H NMR (CD_3CN, 400 MHz): δ 8.23-8.18 (m, 2H), 7.60-7.55 (m, 2H), 7.38-7.29 (m, 5H), 6.04 (br t, NH), 5.26 (s, 2H), 5.09 (s, 2H), 3.94 (d, *J* = 6.2 Hz, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 169.7, 156.3, 147.8, 142.3, 136.0, 128.5, 128.4, 128.2, 123.8, 67.2, 65.5, 42.7. Physical data were in accordance with those reported in literature.[25]

N^{α}-Cbz-Glycine benzyl ester (Z-Gly-OBn)



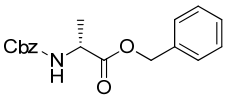
General procedure C using Z-Gly-OH (418 mg, 2.0 mmol), K_2CO_3 (276 mg, 2.0 mmol) and benzyl bromide (262 µl, 2.2 mmol) in DMF (1.5 mL). The product was obtained as a colourless oil after purification by column chromatography (EtOAc/heptane 1:10→1:2) (503 mg, 84%). R_f 0.47 (EtOAc/heptane 1:2). IR (film) 3412, 3027, 2950, 1748, 1717, 1520, 1260, 1189, 1053, 697 cm^{-1}. ^1H NMR (CD_3CN, 300 MHz): δ 7.42-7.28 (m, 10H), 5.96 (br s, NH), 5.15 (s, 2H), 5.08 (s, 2H), 3.89 (d, *J* = 6.3 Hz, 2H). ^{13}C NMR (CD_3CN, 75 MHz): δ 171.0, 157.5, 138.0, 137.0, 129.4, 129.4, 129.1, 129.0, 128.9, 128.7, 67.3, 67.2, 43.3. HRMS (ESI) *m/z* calcd for $C_{17}H_{18}NO_4$ $(M+H)^+$: 300.1236, found: 300.1238. Spectral data were in accordance with those reported in literature.[26]

N^{α}-Cbz-L-Alanine benzyl ester (Z-L-Ala-OBn)



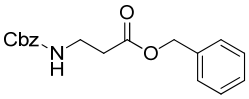
General procedure C using Z-L-Ala-OH (1.12 g, 5.00 mmol), K_2CO_3 (691 mg, 5.00 mmol) and benzyl bromide (654 µl, 5.50 mmol) in DMF (4 mL). The product was obtained as a colourless oil after purification by column chromatography (EtOAc/heptane 1:9→1:2) (1.40 g, 89%). R_f 0.44 (EtOAc/heptane 1:2). $[\alpha]_D^{20}$ −6.0 (*c* 1.02, CH_2Cl_2). IR (film) 3343, 2958, 1719, 1525, 1454, 1257, 1211, 1173, 1069 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.41-7.25 (m, 10H), 5.39 (d, *J* = 6.3 Hz, NH), 5.16 (s, 2H), 5.09 (s, 2H), 4.50-4.36 (m, 1H), 1.40 (d, *J* = 7.2 Hz, 3H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.7, 155.5, 136.2, 135.3, 128.5, 128.4, 128.3, 128.1, 67.0, 66.8, 49.6, 18.5. HRMS (ESI) *m/z* calcd for $C_{18}H_{19}NNaO_4$ $(M+Na)^+$: 336.1212, found: 336.1207. Spectral data were in accordance with those reported in literature.[25]

N^{α}-Cbz-D-Alanine benzyl ester (Z-D-Ala-OBn)



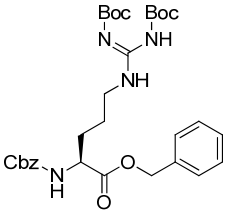
General procedure C using Z-D-Ala-OH (558 mg, 2.50 mmol), K_2CO_3 (346 mg, 2.50 mmol) and benzyl bromide (325 µl, 2.75 mmol) in DMF (2 mL). The product was obtained as a colourless oil after purification by column chromatography (EtOAc/heptane 1:9→1:2) (758 mg, 97%). R_f 0.44 (EtOAc/heptane 1:2). $[\alpha]_D^{20}$ +5.9 (*c* 1.08, CH_2Cl_2). IR (film) 3334, 3027, 1714, 1524, 1454, 1255, 1209, 1172, 1069 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.41-7.25 (m, 10H), 5.37 (d, *J* = 7.0 Hz, NH), 5.16 (s, 2H), 5.10 (s, 2H), 4.50-4.36 (m, 1H), 1.41 (d, *J* = 7.2 Hz, 3H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.6, 155.5, 136.2, 135.3, 128.5, 128.4, 128.4, 128.1, 67.1, 66.9, 49.7, 18.6. HRMS (ESI) *m/z* calcd for $C_{18}H_{20}NO_4$ $(M+H)^+$: 314.1392, found: 314.1385. Spectral data were in accordance with those reported in literature.[27]

N^{α}-Cbz-β-Alanine benzyl ester (Z-β-Ala-OBn)



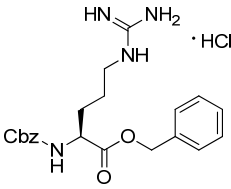
General procedure C using Z-β-Ala-OH (1.12 g, 5.00 mmol), K_2CO_3 (691 mg, 5.00 mmol) and benzyl bromide (654 µl, 5.50 mmol) in DMF (4 mL). The product was obtained as a colourless oil after purification by column chromatography (EtOAc/heptane 1:9→1:2) (1.46 g, 93%). R_f 0.36 (EtOAc/heptane 1:2). IR (film) 3356, 3032, 1716, 1519, 1454, 1245, 1173, cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.40-7.25 (m, 10H), 5.31 (br s, NH), 5.11 (s, 2H), 5.08 (s,2H), 3.47 (dt, *J* = 6.1, 6.1 Hz, 2H), 2.60 (t, *J* = 6.0 Hz, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.0, 156.2, 136.4, 135.5, 128.5, 128.4, 128.3, 128.1, 128.0, 66.6, 66.4, 36.5, 34.4. HRMS (ESI) *m/z* calcd for $C_{18}H_{19}NNaO_4$ $(M+Na)^+$: 336.1212, found: 336.1202. Spectral data were in accordance with those reported in literature.[28]

*N*α-Cbz-L-Arginine(diBoc) benzyl ester (40R)



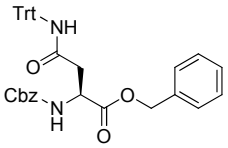
General procedure C using Z-L-Arg(diBoc)-OH (509 mg, 1.00 mmol), K_2CO_3 (138 mg, 1.00 mmol) and benzyl bromide (131 µl, 1.10 mmol). The product was obtained as a colourless oil after purification by column chromatography (EtOAc/heptane 1:3→1:2) (548 mg, 92%). R_f 0.25 (EtOAc/heptane 1:3). $[\alpha]_D^{20}$ –9.1 (*c* 2.74, MeOH). IR (film) 3328, 2958, 2924, 1718, 1637, 1615, 1365, 1329, 1334, 1154, 1132, 1050, 1025 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 11.47 (s, NH), 8.29 (t, *J* = 4.9 Hz, NH), 7.39-7.29 (m, 10H), 5.52 (d, *J* = 8.0 Hz, NH), 5.19 (d, *J* = 12.4 Hz, 1H), 5.14 (d, *J* = 12.4 Hz, 1H), 5.11 (s, 2H), 4.48-4.38 (m, 1H), 3.49-3.27 (m, 2H), 1.95-1.52 (m, 4H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.0, 163.5, 156.2, 155.9, 153.3, 136.2, 135.2, 128.6, 128.5, 128.3, 128.2, 128.1, 83.2, 79.3, 67.3, 67.0, 53.8, 40.2, 29.7, 28.3, 28.1, 25.2. HRMS (ESI) *m/z* calcd for $C_{31}H_{43}N_4O_8$ $(M+H)^+$: 599.3081, found: 599.3075.

*N*α-Cbz-L-Arginine benzyl ester (Z-L-Arg-OBn)



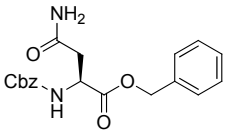
General procedure B1 using **40R** (183 mg, 0.306 mmol), The product was obtained as a colourless oil (127 mg, 95%). $[\alpha]_D^{20}$ –18.3 (*c* 0.95, MeOH). IR (film) 3330, 3162, 2954, 2917, 1701, 1664, 1528, 1263, 1189, 1053 cm^{-1}. ^{1}H NMR (CD_3CN, 300 MHz): δ 7.80 (t, *J* = 4.9 Hz, NH), 7.40-7.25 (m, 10H), 6.83 (br s, 3NH), 6.60 (d, *J* = 7.6 Hz, NH), 5.12 (s, 2H), 5.05 (s, 2H), 4.24-4.12 (m, 1H), 3.15-3.04 (m, 2H), 1.87-1.67 (m, 2H), 1.67-1.54 (s, 2H). ^{13}C NMR (CD_3CN, 75 MHz): δ 173.1, 158.5, 157.4, 138.0, 136.9, 129.5, 129.4, 129.1, 128.9, 128.8, 128.5, 67.5, 67.1, 54.9, 41.6, 29.1, 25.8. HRMS (ESI) *m/z* calcd for $C_{21}H_{27}N_4O_4$ $(M+H)^+$: 399.2032, found: 399.2045.

*N*α-Cbz-L-Asparagine(Trt) benzyl ester (40N)



General procedure C using Z-L-Asn(Trt)-OH (509 mg, 1.00 mmol), K_2CO_3 (138 mg, 1.00 mmol) and benzyl bromide (131 µl, 1.10 mmol). The product was obtained as an off-white powder after purification by column chromatography (EtOAc/heptane 1:3→1:2) (481 mg, 82%). R_f 0.35 (EtOAc/heptane 1:3). Mp 133.7 °C. $[\alpha]_D^{20}$ +2.2 (*c* 2.28, CH_2Cl_2). IR (film) 3325, 3059, 3031, 1722, 1671, 1493, 1448, 1338, 1282, 1256, 1212, 1044 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.34-7.22 (m, 20H), 7.16-7.10 (m, 5H), 6.67 (s, NH), 6.06 (d, *J* = 8.4 Hz, NH), 5.13 (d, *J* = 12.3 Hz, 1H), 5.09 (d, *J* = 12.3 Hz, 1H), 5.09 (s, 2H) 4.64-4.57 (m, 1H), 3.08 (dd, *J* = 4.5, 15.6, 1H), 2.84 (dd, *J* = 4.2, 15.6, 1H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 170.8, 169.0, 156.2, 144.3, 136.3, 135.3, 128.6, 128.5, 128.4, 128.2, 128.0, 127.8, 127.2, 70.9, 67.4, 66.9, 51.1, 38.6. HRMS (ESI) *m/z* calcd for $C_{38}H_{34}Na_1N_2O_5$ $(M+Na)^+$: 621.2365, found: 621.2405.

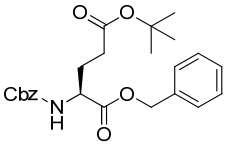
*N*α-Cbz-L-Asparagine benzyl ester (Z-L-Asn-OBn)



General procedure B2 using **40N** (144 mg, 0.241 mmol) and additionally triethylsilane (40 µL, 6.49 mmol) was added. After stirring for 1 hour, the product was obtained as an off-white TFA salt after extraction with Et2O (3 x 20 mL) from water/dioxane (1:2 v/v), washed with heptane and filtrated (70 mg, 61%). R_f 0.40 (EtOAc/heptane 1:4). Mp 126.5 °C. $[\alpha]_D^{20}$ –13.6 (*c* 2.15, MeOH). IR (film) 3404, 3312, 3205, 3067, 3032, 1742, 1696, 1657, 1542, 1339, 1258, 1218, 1059 cm^{-1}. ^{1}H NMR (CD_3CN, 400 MHz): δ 7.39-7.29 (m, 10H), 6.26-6.16 (m, 2NH), 5.75 (br s, NH), 5.13 (s, 2H), 5.09 (d, *J* = 12.8 Hz, 1H), 5.06 (d, *J* = 12.4 Hz, 1H), 4.55 (dt, *J* = 5.7, 8.8 Hz, 1H), 2.75 (dd, *J* = 6.0, 16.1 Hz, 1H), 2.69 (dd, *J* = 5.0, 16.0 Hz, 1H). ^{13}C NMR (CD_3CN, 75 MHz): δ 172.2, 157.0, 145.0, 130.1,

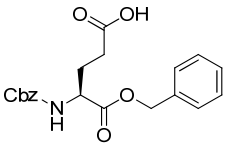
129.4, 129.3, 129.0, 128.9, 128.8, 128.6, 127.3, 67.6, 67.2, 51.8, 37.3. HRMS (ESI) *m/z* calcd for C19H20Na1N2O5 (M+Na)+: 379.1270, found: 379.1271.

N^{α}-Cbz-L-Glutamate(O^{t}Bu) benzyl ester (40E)



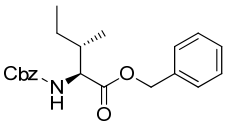
General procedure C using Z-L-Glu(O^{t}Bu)-OH (337 mg, 1.00 mmol), K_2CO_3 (138 mg, 1.00 mmol) and benzyl bromide (131 µl, 1.10 mmol). The product was obtained as an off-white powder after purification by column chromatography (EtOAc/heptane 1:3→1:2) (363 mg, 85%). R_f 0.35 (EtOAc/heptane 1:3). Mp 47.5 °C. $[\alpha]_D^{20}$ –1.3 (*c* 2.39, CH_2Cl_2). IR (film) 3339, 3028, 2976, 2885, 1722, 1525, 1455, 1367, 1256, 1212, 1151, 1056 cm^{-1}. ^{1}H NMR ($CDCl_3$, 400 MHz): δ 7.39-7.29 (m, 10H), 5.43 (d, *J* = 8.2 Hz, NH), 5.17 (s, 2H), 5.10 (s, 2H), 4.48-4.40 (m, 1H), 2.36-2.10 (m, 3H), 2.01-1.90 (m, 1H), 1.42 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.9, 171.8, 155.9, 136.2, 135.2, 128.6, 128.5, 128.3, 128.1, 128.0, 80.8, 67.3, 67.0, 53.5, 31.3, 28.0, 27.6. HRMS (ESI) *m/z* calcd for $C_{24}H_{30}N_1O_6$ (M+H)$^+$: 428.2073, found: 428.2067. Spectral data were in accordance with those reported in literature.[29]

N^{α}-Cbz-L-Glutamate benzyl ester (Z-L-Glu-OBn)



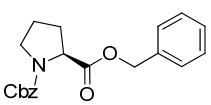
General procedure B2 using **40E** (106 mg, 0.234 mmol) and additionally Et_3SiH (40 µL, 6.49 mmol) was added. After stirring for 1 hour, the product was obtained as a colourless oil (89 mg, 78%). $[\alpha]_D^{20}$ –21.4 (*c* 1.54, MeOH). IR (film) 3327, 3057, 3033, 2956, 1706, 1525, 1454, 1343, 1256, 1211, 1154, 739, 697 cm^{-1}. ^{1}H NMR (CD_3CN, 300 MHz): δ 7.39-7.29 (m, 10H), 6.04 (d, *J* = 7.5 Hz, NH), 5.14 (s, 2H), 5.07 (s, 2H), 4.30-4.21 (m, 1H), 2.37 (t, *J* = 7.3 Hz, 2H), 2.15-2.03 (m, 1H), 1.95-1.83 (m, 1H). ^{13}C NMR (CD_3CN, 75 MHz): δ 174.1, 172.7, 157.1, 138.0, 137.0, 129.5, 129.4, 129.1, 129.0, 128.9, 128.6, 67.6, 67.2, 54.5, 30.2, 27.3. HRMS (ESI) *m/z* calcd for $C_{20}H_{21}Na_1N_1O_6$ (M+Na)$^+$: 394.1267, found: 394.1269. Spectral data were in accordance with those reported in literature.[30]

N^{α}-Cbz-L-Isoleucine benzyl ester (Z-L-Ile-OBn)



General procedure C using Z-L-Ile-OH (1.33 g, 5.00 mmol), K_2CO_3 (691 mg, 5.00 mmol) and benzyl bromide (654 µl, 5.50 mmol) in DMF (4 mL). The product was obtained as a colourless oil after purification by column chromatography (EtOAc/heptane 1:9→1:2) (1.63 g, 91%). R_f 0.56 (EtOAc/heptane 1:2). $[\alpha]_D^{20}$ +1.5 (*c* 1.02, CH_2Cl_2). IR (film) 3343, 3067, 1716, 1518, 1455, 1335, 1192 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 7.48-7.24 (m, 10H), 5.35 (d, *J* = 7.1 Hz, NH), 5.19 (d, *J* = 12.1 Hz, 1H), 5.12 (d, *J* = 12.1 Hz, 1H), 5.09 (s, 2H), 4.46-433 (m, 1H), 1.99-1.73 (m, 1H), 1.46-1.23 (m, 1H), 1.23-1.00 (m, 1H), 1.00-0.74 (m, 6H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 171.8, 156.0, 136.2, 135.2, 128.5, 128.4, 128.3, 128,2, 128.0, 128.0, 66.9, 58.3, 37.9, 24.8, 15.4, 11.5. HRMS (ESI) *m/z* calcd for $C_{21}H_{26}NO_4$ (M+H)$^+$: 356.1862, found: 356.1863.

N^{α}-Cbz-L-Proline benzyl ester (Z-L-Pro-OBn)



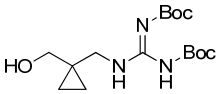
General procedure C using Z-L-Pro-OH (1.25 g, 5.00 mmol), K_2CO_3 (691 mg, 5.00 mmol) and benzyl bromide (654 µl, 5.50 mmol) in DMF (4 mL). The product was obtained as a colourless oil after purification by column chromatography (EtOAc/heptane 1:9→1:2) (1.66 g, 98%). R_f 0.36 (EtOAc/heptane 1:2). $[\alpha]_D^{20}$ –47.3 (*c* 2.01, CH_2Cl_2). IR (film) 3062, 3032, 1744, 1703, 1415, 1350, 1166, 1118 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 7.39-7.18 (m, 10H), 5.24-4.94 (m, 4H), 4.44 (dd, *J* = 3.4, 8.6 Hz, 0.45H), 4.36 (dd, *J* = 3.7, 8.6 Hz, 0.55H), 3.68-3.40 (m, 2H), 2.28-2.09 (m, 1H), 2.06-1.76 (m, 3H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 172.4, 172.2, 154.6, 154.0, 136.5, 136.3, 135.5, 135.3, 128.3, 128.2, 128.1, 128.0, 127.9, 127.8, 127.8, 127.7, 127.6,

127.6, 127.5, 66.7, 66.7, 66.5, 66.4, 59.1, 58.7, 46.7, 46.2, 30.6, 29.6, 24.0, 23.3. HRMS (ESI) *m/z* calcd for $C_{20}H_{21}NNaO_4$ $(M+Na)^+$: 362.1368, found: 362.1364. Spectral data were in accordance with those reported in literature.[31]

3-Amino-2-cyclopropylpropanol (42)

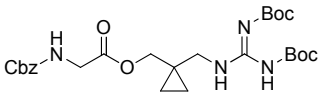
To a solution of $LiAlH_4$ (244 mg, 6.43 mmol, 6.4 equiv) in anhydrous THF (5 mL) containing 4 Å MS was added 1-(aminocarbonyl)cyclopropanecarboxylic acid (129 mg, 1.0 mmol, 1.0 equiv) and the mixture was refluxed overnight. Work-up was performed by evaporation of the solvent and slow addition of water (5 mL). Subsequently, THF (20 mL) was added and stirred for 30 minutes The resulting reaction mixture was evaporated *in vacuo*, and Et_2O (50 mL) was added to form a precipitate, which was filtered off. The residue was washed with Et_2O (3 × 50 mL) and the organic layers were combined, dried over Na_2SO_4, filtered, and evaporated *in vacuo* to afford **42** (74 mg, 55%) as a yellow oil. R_f 0.16 (10% MeOH in CH_2Cl_2). IR (film) 3356, 3280, 2996, 2920, 2865, 1653, 1598, 1425, 1023, 988, 933 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 3.62 (s, 2H), 2.82 (s, 2H), 2.48 (br s, 2NH, 1OH), 0.51-0.46 (m, 2H), 0.42-0.37 (m, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 68.2, 49.7, 30.9, 10.1, 9.8. HRMS (ESI) *m/z* calcd for $C_5H_{12}NO$ $(M+H)^+$: 102.0919, found: 102.0902.

3-[*N,N'*-di(Boc)guanidino]-2-cyclopropylpropanol (43)



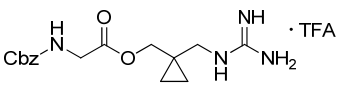
To a solution of *N,N'*-di(Boc)-*S*-methylisothiourea (290 mg, 1.0 mmol, 1.0 equiv) and DMAP (183 mg, 1.5 mmol, 1.5 equiv) in CH_2Cl_2 (5 mL) was added **42** (0.101 g, 1.0 mmol, 1.0 equiv) and stirred at room temperature for 5 hours. The reaction mixture was evaporated to dryness and purified by means of column chromatography (MeOH in CH_2Cl_2, 0→5%) to afford **43** (341 mg, 99%) as a white fluffy solid. R_f 0.32 (5% MeOH in CH_2Cl_2). IR (film) 3328, 2968, 2920, 1722, 1660, 1612, 1556, 1425, 1369, 1314, 1238, 1162, 1134, 1058, 815, 753 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.43 (br s, NH), 8.57 (br s, NH), 5.03 (t, *J* = 7.5 Hz, 1H, OH), 3.35-3.29 (m, 4H), 1.51 (s, 9H), 1.48 (s, 9H), 0.50-0.45 (m, 4H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 162.8, 157.0, 153.1, 83.5, 79.5, 66.0, 45.6, 30.9, 28.1, 23.4, 9.3. HRMS (ESI) *m/z* calcd for $C_{16}H_{30}N_3O_5$ $(M+H)^+$: 344.2186, found: 344.2172.

N^{α}-Cbz-Glycine 3-[*N,N'*-di(Boc)guanidino]-2-cyclopropylpropyl ester (44)



General procedure A1 with adjusted amounts of every component using **43** (100 mg, 0.29 mmol, 1.0 equiv), DCC (72 mg, 0.35 mmol, 1.2 equiv), DMAP (43 mg, 0.35 mmol, 1.2 equiv) and Z-Gly-OH (73 mg, 0.35 mmol, 1.2 equiv) in CH_2Cl_2 (2 mL). The product (**44**) was obtained as a white solid after purification by column chromatography (MeOH in CH_2Cl_2, 0→5%) (155 mg, quant). R_f 0.44 (5% MeOH in CH_2Cl_2). IR (film) 3315, 2982, 1715, 1639, 1605, 1563, 1425, 1362, 1328, 1168, 1134, 1058, 815, 746, 691 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.52 (s, 1H, NH), 8.59 (s, 1H, NH), 7.36-7.33 (m, 5H), 5.57 (s, 1H, NH), 5.13 (s, 2H), 4.09-4.06 (m, 4H), 3.41 (s, 2H), 1.49 (s, 18H), 0.88-0.62 (m, 4H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 206.8, 170.1, 163.4, 156.3, 156.0, 153.2, 136.3, 128.4, 128.2, 128.2, 128.1, 83.2, 71.4, 67.2, 66.9, 47.3, 43.1, 29.3, 18.9, 10.8. HRMS (ESI) *m/z* calcd for $C_{26}H_{39}N_4O_8$ $(M+H)^+$: 535.2768, found: 535.2747.

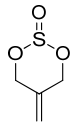
N^{α}-Cbz-Glycine 3-guanidino-2-cyclopropylpropyl ester (Z-Gly-O3G∇)



Compound **44** (65 mg, 0.12 mmol) was treated with CH_2Cl_2/TFA (2 mL, 1:1) overnight, the solvents were removed under reduced pressure and co-evaporated with ^{t}BuOH (3 × 10 mL). **Z-Gly-O3G∆** (55 mg, quant) was obtained as a colourless oil. IR (film) 3349, 3169, 1674, 1535, 1258, 1189, 1134, 1037, 988, 836,

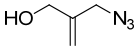
788, 760, 718, 712 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.35-7.33 (m, 5H), 7.26 (br s, 2NH), 6.84 (br s, 3NH), 5.76 (br s, NH), 5.10 (s, 2H), 4.02 (s, 2H), 3.90 (s, 2H), 3.03 (s, 2H), 0.61-0.57 (m, 4H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 170.0, 157.5, 157.2, 135.8, 128.6, 128.4, 127.8, 70.1, 67.5, 46.9, 43.1, 31.0, 19.4, 9.8. HRMS (ESI) *m/z* calcd for $C_{16}H_{23}N_4O_4$ $(M+H)^+$: 335.1719, found: 335.1713.

5-Methylene-2-oxo[1,3,2]dioxathiane (46)



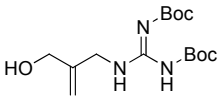
To an emulsion of diol 2-methylene-1,3-propanediol (1.76 g, 20 mmol, 1.0 equiv) in CCl_4 (12 mL) at 0 °C was added dropwise a solution of thionyl chloride (3.53 g, 30 mmol, 1.5 equiv) in CCl_4 (6 mL) under vigorous stirring. When the evolution of HCl had ceased after approximately 30 minutes, the solution was stirred for an additional 15 minutes Evaporation of the solvent at 0 °C and 5 mbar, followed by Kugelrohr distillation (b.p. 90-110 °C/ >10 Torr) yielded **46** (1.95 g, 73 % yield) as a colourless liquid. R_f0.55 (EtOAc/pentane 1:4). IR (film) 3088, 2991, 2938, 2873, 1461, 1446, 1416, 1342, 1300, 1239, 1195, 1177, 982, 958, 926, 870, 759, 717, 692, 663 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 5.35 (dt, *J* = 1.5, 13.2 Hz, 2H), 5.13-5.11 (m, 2H), 4.24 (dt, *J* =1.5, 13.2 Hz, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 135.5, 114.4, 61.6. Spectral data were in accordance with those reported in literature.[7]

3-Azido-2-methylenepropanol (47)



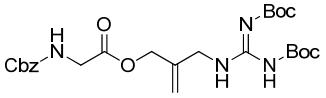
To a solution of **46** (630 mg, 4.7 mmol, 1.0 equiv) in DMF (10 mL) was added sodium azide (366 mg, 5.6 mmol, 1.2 equiv) and the mixture was heated to 80 °C. After 30 minutes, the reaction was quenched by addition of water (20 mL) and extracted with Et_2O (3 × 10 mL). The combined organic extracts were washed with water (20 mL), dried with $MgSO_4$ and concentrated *in vacuo* at 0 °C and 5 mbar. The crude product was purified by column chromatography (EtOAc/pentane 1:4) to obtain **47** (478 mg, 90%) as a yellow oil. R_f0.24 (EtOAc/pentane 1:4). IR (film) 3308, 2913, 2858, 2089, 1653, 1445, 1238, 1072, 1016, 919, 871, 656, 559 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 5.28 (s, 1H), 5.18 (s, 1H), 4.19 (s, 2H), 3.86 (s, 2H), 1.72 (s, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 142.7, 114.6, 64.0, 53.3. Spectral data were in accordance with those reported in literature.[7]

3-[*N,N'*-di(Boc)guanidino]-2-methylenepropanol (49)



To a solution of **47** (113 mg, 1.0 mmol, 1.1 equiv) in water/THF (30 mL, 1:5) was added trimethylphosphine (84 mg, 1.1 mmol, 1.1 equiv) and the resulting mixture was stirred at room temperature overnight. Subsequently, HCl (1.5 mL, 1.0 M) was added and the solvent was evaporated *in vacuo* to afford crude **48** as a yellow oil. Without further purification crude **48** was added to a solution of *N,N'*-di(Boc)-*S*-methylisothiourea (145 mg, 0.50 mmol) and DMAP (92 mg, 0.75 mmol) in CH_2Cl_2 (5 mL) and the mixture was stirred at room temperature for 5 h. After evaporation to dryness and purification by means of column chromatography (MeOH in CH_2Cl_2, 0→5%) **49** was obtained as a white fluffy solid (242 mg, 74% over two steps). R_f 0.63 (5% MeOH in CH_2Cl_2). IR (film) 3324, 2952, 2927, 1729, 1653, 1612, 1432, 1301, 1245, 1167, 1127, 1085, 760, 680, 615 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.43 (br s, NH), 8.57 (br s, NH), 5.05 (s, 1H), 4.99 (s, 1H), 4.87 (br s, OH), 4.08-4.04 (m, 4H), 1.50 (s, 9H), 1.48 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 162.6, 156.7, 153.1, 145.8, 114.0, 83.6, 79.6, 63.4, 43.0, 28.2, 28.0.

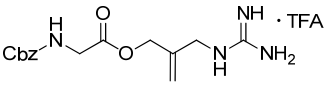
N^{α}-Cbz-Glycine 3-[*N,N'*-di(Boc)guanidino]-2-methylenepropyl ester (50)



General procedure A1 with adjusted amounts of every component using **49** (120 mg, 0.40 mmol, 1.0 equiv), DCC (83 mg, 0.40 mmol, 1.0 equiv), DMAP (49 mg, 0.40 mmol, 1.0 equiv) and Z-Gly-OH (84 mg, 0.40 mmol, 1.0 equiv) in EtOAc

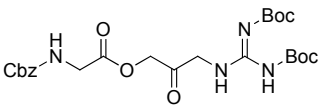
(2 mL). Purification was performed by column chromatography (EtOAc/heptane 1:3) to afford **50** as a white solid (181 mg, 95%). R_f 0.41 (EtOAc/heptane 1:1). IR (film) 3328, 2982, 2927, 1729, 1653, 1612, 1411, 1369, 1321, 1245, 1168, 1127, 1092, 1072, 822, 760, 691, 615 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.52 (br s, NH), 8.48 (br s, NH), 7.41-7.32 (m, 5H), 5.45 (br s, NH), 5.24 (s, 2H), 5.13 (s, 2H), 4.66 (s, 2H), 4.16-4.10 (m, 2H), 4.09-4.00 (m, 2H), 1.49 (s, 18H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 163.4, 156.2, 153.3, 139.1, 136.3, 128.5, 128.2, 128.1, 117.1, 83.4, 79.5, 67.0, 66.4, 43.2, 42.9, 34.0, 28.2, 28.1, 24.9. HRMS (ESI) *m/z* calcd for $C_{25}H_{37}N_4O_8$ $(M+H)^+$: 521.2611, found: 521.2592.

N^{α}-Cbz-Glycine 3-guanidino-2-methylenepropyl ester (Z-Gly-O3G=)



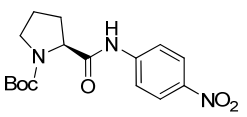
Compound **50** (100 mg, 0.20 mmol) was treated with CH_2Cl_2/TFA (6 mL, 1:1) overnight, the solvents were removed under reduced pressure and co-evaporated with ^{t}BuOH (3 × 10 mL). **Z-Gly-O3G=** was obtained as a colourless oil (83 mg, quant). IR (film) 3356, 3190, 2941, 1667, 1528, 1452, 1272, 1182, 1044, 975, 836, 795, 732, 691 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 7.67 (br s, NH), 7.34-7.32 (m, 5H), 6.91 (br s, 2NH), 5.59 (br s, NH), 5.33 (d, *J* = 11.1 Hz, 2H), 5.10 (s, 2H), 4.67 (s, 2H), 3.92 (d, *J* = 6.0 Hz, 2H), 3.77 (d, *J* = 6.3 Hz, 2H), 3.38 (br s, 2H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 169.7, 157.6, 137.8, 135.8, 128.6, 128.4, 127.8, 118.9, 67.5, 66.0, 43.9, 43.0, 33.2, 31.2. HRMS (ESI) *m/z* calcd for $C_{15}H_{21}N_4O_4$ $(M+H)^+$: 321.1563, found: 321.1555.

N^{α}-Cbz-Glycine-3-[*N,N'*-di(Boc)guanidino]-2-carbonylpropyl ester (51)



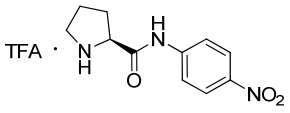
A solution of **50** (260 mg, 0.50 mmol, 1.0 equiv) in CH_2Cl_2 (10 mL) was stirred and cooled to −78 °C. Subsequently, ozone was bubbled through the solution for 30 minutes. The solution turned light blue and TLC-analysis indicated the disappearence of the substrate. Then DMS (34 mg, 0.55 mmol, 1.1 equiv) was added in one portion, stirring still continued at −78 °C for 10 minutes. The reaction mixture was brought to room temperature and stirred for an additional 10 minutes. Then the reaction was quenched with water (10 mL), the layers were separated and the waterlayer was washed with CH_2Cl_2 (3 × 5 mL). The combined organic layers were dried with Na_2SO_4 and the solvent was evaporated *in vacuo*. Purification was performed by means of column chromatography (EtOAc/heptane 1:4) to afford **51** (159 mg, 61%). R_f 0.50 (EtOAc/heptane 1:1). IR (film) 3322, 2975, 2920, 1715, 1646, 1618, 1563, 1411, 1355, 1300, 1155, 1065, 753 cm^{-1}. ^1H NMR ($CDCl_3$, 300 MHz): δ 11.39 (br s, NH), 8.96 (br s, NH), 7.36-7.32 (m, 5H), 5.52 (br s, NH), 5.13 (s, 2H), 4.81 (s, 2H), 4.36 (s, 2H), 4.12 (d, *J* = 5.7 Hz, 2H), 1.50 (s, 18H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 209.1, 170.4, 157.6, 136.8, 136.4, 128.6, 128.4, 127.8, 115.9, 67.5, 67.1, 66.0, 42.8, 30.1.

N^{α}-Boc-L-Proline-*p*-nitroanilide (Boc-Pro-*p*NA)



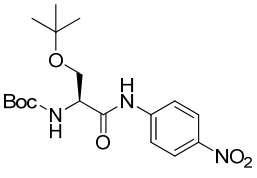
General procedure D using Boc-Pro-OH (2.15 g, 10 mmol), *p*-nitroaniline (1.38 g, 10 mmol) and phosphorus oxychloride (1.01 ml, 11 mmol). After recrystallisation with isopropyl alcohol and heptane the product was obtained as a slightly yellowish solid (2.19 g, 65%). ^1H NMR ($CDCl_3$, 300 MHz): δ 10.24 (br s, NH), 8.23-8.14 (m, 2H), 7.72-7.63 (m, 2H), 4.49 (br s, NH), 3.52-3.26 (m, 2H), 2.66-2.47 (m, 1H), 2.04-1.89 (m, 3H), 1.51 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 170.6, 156.7, 144.3, 143.1, 124.9, 118.9, 81.4, 60.6, 47.4, 28.4, 27.3, 24.5. HRMS (ESI) *m/z* calcd for $C_{16}H_{21}N_3NaO_5$ $(M+Na)^+$: 358.1379, found: 358.1377. Spectral data were in accordance with those reported in literature.[32]

L-Proline-*p*-nitroanilide (Pro-*p*NA)



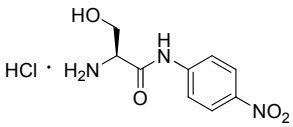
*N*α-Boc-L-Proline-*p*-nitroanilide (700 mg, 2.09 mmol, 1 equiv) was dissolved in CH_2Cl_2 (10 mL) to which TFA was added (2.5 mL). The reaction mixture was stirred at room temperature overnight. The solvents were removed under reduced pressure and co-evaporated with CH_2Cl_2 (3 × 20 mL). The resulting solid was purified by recrystallisation with EtOAc and heptane (652 mg, 94%). ^{1}H NMR (CD_3CN, 300 MHz): δ 10.14 (s, NH), 8.24-8.16 (m, 2H), 7.84-7.77 (m, 2H), 4.65-4.56 (m, 1H), 3.48-3.31 (m, 2H), 2.54-2.40 (m, 2H), 2.10-1.95 (m, 2H). ^{13}C NMR (CD_3CN, 75 MHz): δ 168.7, 144.8, 141.2, 125.8, 120.4, 118.2, 61.3, 47.5, 30.4, 25.1. HRMS (ESI) *m/z* calcd for $C_{11}H_{14}N_3O_3$ $(M+H)^+$: 236.1035, found: 236.1035. Spectral data were in accordance with those reported in literature.[33]

***N*α-Boc-L-Serine(O*t*Bu)-*p*-nitroanilide (Boc-Ser(O*t*Bu)-*p*NA)**



General procedure D using Boc-Ser(O*t*Bu)-OH (862 mg, 3.3 mmol), *p*-nitroaniline (456 mg, 3.3 mmol) and phosphorus oxychloride (332 µl, 3.6 mmol). After purification by column chromatography (EtOAc/heptane 1:10→1:1) the product was obtained as a slightly yellowish solid (573 mg, 46%). R_f 0.49 (EtOAc/heptane 1:1). Mp 55 °C. $[\alpha]_D^{20}$ –24.8 (*c* 0.75, CH_2Cl_2). IR (film) 3296, 2976, 2928, 1679, 1511, 1341, 1162 cm^{-1}. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 9.14 (s, NH), 8.26-8.17 (m, 2H), 7.72-7.63 (m, 2H), 5.53-5.41 (m, NH), 4.41-4.27 (m, 1H), 3.90 (dd, *J* = 3.9, 8.6 Hz, 1H), 3.48 (dd, *J* = 8.4, 8.4 Hz), 1.48 (s, 9H), 1.26 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz): δ 169.5, 155.6, 143.6, 143.4, 125.1, 119.0, 80.7, 74.9, 61.6, 54.9, 28.3, 27.5. HRMS (ESI) *m/z* calcd for $C_{18}H_{28}N_3O_6$ $(M+H)^+$: 382.1978, found: 382.1978.

L-Serine-*p*-nitroanilide (Ser-*p*NA)



*N*α-Boc-L-Serine(O*t*Bu)-*p*-nitroanilide (500 mg, 1.31 mmol, 1 equiv) was treated with a commercially available solution of ~ 4 M HCl in dioxane (10 mL) and stirred at room temperature overnight. The solvents were removed under reduced pressure and co-evaporated with n-BuOH (3 × 20 mL). The resulting solid was purified by recrystallisation with isopropyl alcohol and heptane (290 mg, 85%). R_f 0.16 (10% MeOH in CH_2Cl_2). Mp 170 °C. $[\alpha]_D^{20}$ +13.8 (*c* 0.33, MeOH). IR (film) 3204, 3047, 1702, 1569, 1506, 1351, 1261 cm^{-1}. ^{1}H NMR (DMSO-d_6, 300 MHz): δ 8.30-8.26 (m, 2H), 7.95-7.88 (m, 2H), 5.59 (t, *J* = 5.0 Hz, NH), 4.14 (t, *J* = 4.4 Hz, 1H), 4.00-3.80 (m, 2H). ^{13}C NMR (DMSO-d_6, 75 MHz): δ 166.7, 144.3, 142.6, 124.9, 119.1, 60.0, 55.1. HRMS (ESI) *m/z* calcd for $C_9H_{12}N_3O_4$ $(M+H)^+$: 226.0828, found: 226.0822.

7.11 References

[1] L. Zervas, T. T. Otani, M. Winitz, J. P. Greenstein, *J. Am. Chem. Soc.* **1959**, *81*, 2878-2884.

[2] A. S. Verdini, P. Lucietto, G. Fossati, C. Giordani, *Tetrahedron Lett.* **1992**, *33*, 6541-6542.

[3] (a) A. Isidro-Llobet, M. Álvarez, F. Albericio, *Chem. Rev.* **2009**, *109*, 2455–2504; (b) R. Paul, A. S. Kende, *J. Am. Chem. Soc.* **1964**, *86*, 4162-4166; (c) S. Mojsov, A. R. Mitchell, R. B. Merrifield, *J. Org. Chem.* **1980**, *45*, 555-560.

[4] C. b. M. Schwarz, *Synlett* **2000**, *9*, 1369.

[5] P. Kuhl, U. Zaeharias, H. Burckhardt, H. D. Jakubke, *Monatsh. Chem.* **1986**, *117*, 1195-1204.

[6] M. Buchholz, U. Heiser, S. Schilling, A. J. Niestroj, K. Zunkel, H.-U. Demuth, *J. Med. Chem.* **2005**, *49*, 664-677.
[7] M. Friedrich, A. I. Savchenko, A. Wächtler, A. d. Meijere, *Eur. J. Org. Chem.* **2003**, *11*, 2138-2143.
[8] D. T. S. Rijkers, H. P. H. M. Adams, H. C. Hemker, G. I. Tesser, *Tetrahedron* **1995**, *51*, 11235-11250.
[9] S. Thust, B. Koksch, *J. Org. Chem.* **2003**, *68*, 2290-2296.
[10] F. Rodriguez, I. Rozas, J. E. Ortega, J. J. Meana, L. F. Callado, *J. Med. Chem.* **2007**, *50*, 4516-4527.
[11] T. Yaegashi, S. Nunomura, T. Okutome, T. Nakayama, M. Kurumi, Y. Sakurai, T. Aoyama, S. Fujii, *Chem. Pharm. Bull.* **1984**, *32*, 4466-4477.
[12] V. Rerat, G. Dive, A. A. Cordi, G. C. Tucker, R. Bareille, J. Amédée, L. Bordenave, J. Marchand-Brynaert, *J. Med. Chem.* **2009**, *52*, 7029-7043.
[13] B. J. Neubert, B. B. Snider, *Org. Lett.* **2003**, *5*, 765-768.
[14] D. J. Upadhyaya, A. Barge, R. Stefania, G. Cravotto, *Tetrahedron Lett.* **2007**, *48*, 8318–8322.
[15] A. J. Wilson, M. Masuda, R. P. Sijbesma, E. W. Meijer, *Angew. Chem. Int. Ed.* **2005**, *44*, 2275-2279.
[16] J. A. Hodges, R. T. Raines, *Org. Lett.* **2006**, *8*, 4695-4697.
[17] C. Wiles, P. Watts, S. P. Haswell, *Tetrahedron Lett.* **2006**, *47*, 5261–5264.
[18] T. Tsuji, S. Okusada, M. Maeda, K. Kawasaki, *Chem. Pharm. Bull.* **1986**, *34*, 2214-2217.
[19] (a) S.-S. P. Chou, Y.-H. Yeh, *Tetrahedron Lett.* **2001**, *42*, 1309-1311; (b) M. R. Pitts, J. R. Harrison, C. J. Moody, *J. Chem. Soc. Perkin Trans. 1* **2001**, 955–977.
[20] T. C. Bruice, F.-H. Marquardt, *J. Am. Chem. Soc.* **1962**, *84*, 365-370.
[21] V. K. Chadha, K. G. Leidal, B. V. Plapp, *J. Med. Chem.* **1983**, *26*, 916-922.
[22] J. A. Wendt, H. Stenmark, L. Wu, Y. Wang, B. B. Chen, T. D. Penning, V. Downs, M. L. Boys, M. Russell, D. P. Spangler, *Vol. WO2004/58760 A1, 2004* **2004**.
[23] J. Martinez, J. Laur, B. Castro, *Tetrahedron* **1985**, *41*, 739-743, .
[24] K. Okano, N. Mitsuhashi, H. Tokuyama, *Chem. Commun.* **2010**, *46*, 2641–2643.
[25] T. Miyazawa, K. Tanaka, E. Ensatsu, R. Yanagihari, T. Yamada, *J. Chem. Soc. Perkin Trans. 1* **2001**, 87-93.
[26] T. Iwasaki, Y. Maegawa, Y. Hayashi, T. Ohshima, M. K., *J. Org. Chem.* **2008**, *73*, 5147–5150.
[27] R. Paruszewski, G. Rostafińska-Suchar, M. Strupińska, I. Winiecka, J. P. Stables, *Pharmazie* **2000**, *55*, 27-30.
[28] M. C. Sheikh, S. Takagi, A. Ogasawara, M. Ohira, R. Miyatake, H. Abe, T. Yoshimura, H. Morita, *Tetrahedron* **2010**, *66*, 2132-2140.
[29] E. Barbayianni, I. Fotakopoulou, M. W. Schmidt, V. Constantinou-Kokotou, U. T. Bornscheuer, G. Kokotos, *J. Org. Chem.* **2005**, *70*, 8730-8733.
[30] S. Datta, L. Hoesch, *Helv. Chim. Acta* **1987**, *70*, 1261-1267.
[31] K. Hattori, H. Sajiki, K. Hirota, *Tetrahedron* **2000**, *56*, 8433-8441.
[32] S. V. Pansare, R. L. Kirby, *Tetrahedron* **2009**, *65*, 4557–4561.
[33] J. Moorthy, S. Saha, *Eur. J. Org. Chem.* **2009**, 739–748.

Chapter 8

Perspective

This chapter provides my personal view on the research described in this thesis and the options that I see for chemoenzymatic peptide synthesis based on either substrate mimetics or enzymatic activation in the future.

8.1 Theoretical potential of the substrate mimetics concept

I was charmed by the concept that a hydrolytic enzyme can be misled by a substrate mimetic in such a way that it catalyses synthesis under aqueous circumstances. Substrate specificity is often quite strict, which generally serves a good reason in the native environment. It is remarkable that the addition of a specific element required for recognition to an amino acid of choice, can bring about its acceptance as a substrate. This sounds like an ideal situation for synthetic purposes.

More particularly, the key players in the substrate mimetics approach are the guanidinophenyl ester (the mimetic) and the arginine-specific serine protease trypsin. For a viable industrial process, the OGp ester needs to be simplified and the alternative moiety should preferably be enzymatically coupled to the amino acid or peptide. Whether this should be carried out in separate steps or in a two-enzyme-one-pot process is not yet the subject of debate. If these goals could be realised, industry would be provided with a promising chemoenzymatic peptide synthesis strategy.

8.2 Comprehensive overview of the results described in this thesis

First a proper insight was required into the applicability of the substrate mimetics strategy. Although we wanted to replace the OGp ester in the end, we started out to subject this ester to the cysteine protease papain, in order to potentially widen the scope of the substrate mimetics strategy. Possibly due to the broad substrate specificity of papain, this mimetic – enzyme combination turned out to work well. Various dipeptides were successfully synthesised in aqueous media. A computational docking study, intended to provide insight into the molecular interactions, revealed that the OGp ester binds to the active site of papain in a different manner than the natural substrate arginine. This was reason for us to henceforth refer to enzyme-specific activation, a more comprehensive designation for this phenomenon, which also covers substrate mimetics.

This finding made it more likely to find a simpler alternative for OGp. From a set of potentially activating esters, two compounds appeared to be good replacements. The ODmap esters reacted significantly faster than the OBn esters, but the scope of the dipeptide synthesis reaction remained similar compared to OGp in both cases. To really establish the potential of these alternative esters, two relevant dipeptides were synthesised on gram scale. Z-Gly-Phe-NH_2 starting from Z-Gly-ODmap and H-L-Glu-Trp-OH starting from Z-L-Glu-OBn, which itself was enzymatically prepared with Alcalase-CLEA in anhydrous media, a system developed by DSM. Now full advantage is taken from the enzymatic coupling.

Thorough evaluation of the potentially activating esters gave rise to the question to what extent recognition is actually important and to what extent the leaving group ability of the ester is determining for its activity. Experimental research combined with computational methods such as docking and *ab initio* calculations on trypsin, the most studied enzyme in this context, showed that electronic properties are decisive. Obviously, as a precondition,

the ester has to fit in the active site, and a nice fit certainly contributes to a higher activity, but that alone is insufficient.

In this light it is less surprising that OGp, OTmap, ODmap, OTfe and OCam structurally outrival the other esters, even in enzymes which favour large aromatic or hydrophobic residues in their S_1 subsite such as α-chymotrypsin and Alcalase and even when the reaction is performed in organic solvents. They are simply the most active esters present in the set. The initially remarkable observation that the phenolic esters containing a positive charge give the highest activity, despite the native preference of α-chymotrypsin and Alcalase, may be rationalised by the nature of the interactions that can be made. Ionic interactions are much stronger than for example Van der Waals interactions, so, if it is possible to make such an interaction, this is relatively favourable.

The results obtained from an attempt to design mimetics for the proline protease MaxiPro,[1] not previously described in this thesis, are nicely in keeping with this view too. In Figure 8.1 potential proline mimetics are depicted.

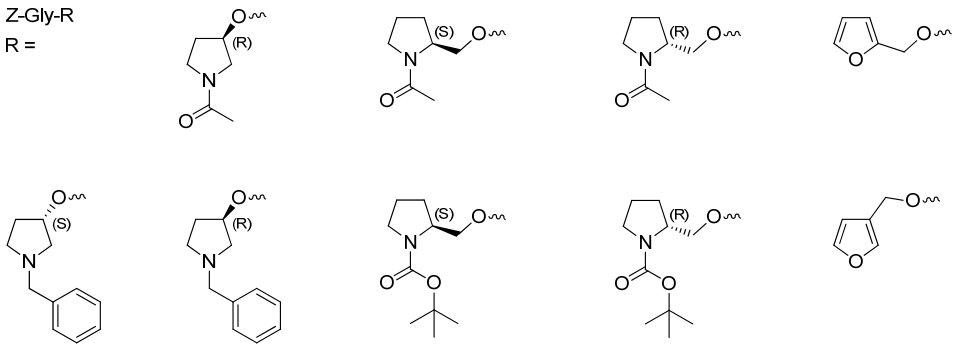


Figure 8.1 *Set of potentially activating esters for proline protease MaxiPro*

Most esters appeared to be slightly active in a dipeptide synthesis assay with H-Phe-NH_2, but no reaction was fast and promising, for none of them is a good leaving group. The set was supplemented with OGp, OCam and OTfe, which were additionally evaluated. Interesting in this respect is the complete inactivity of the OGp ester. As it is presumed that the S_1 subsite of the enzyme is built in such a manner that it specifically disposes proline,[2] probably based on the β-turn that is induced by the presence of this amino acid in a peptide chain, it is likely that OGp does not fit in the active site. Due to the unavailability of a crystal structure of MaxiPro, this hypothesis could not be tested. OTfe and OCam on the other hand performed similarly well as the other esters, indicating that these small esters do fit.

8.3 Bottlenecks encountered throughout the project

In an early stage of the research, lack of insight into the enzymes hindered the progress. The collaboration with a computational modelling group has proven to be extremely

valuable to gain insight into the active site of our enzymes. A working model for papain was built, which at least helped rationalising the functioning of our activating esters, and gave direction to our search. Yet, it remains a model, so a critical attitude is appropriate. A crystal structure of papain co-crystallised with an inactive analogue of an OGp ester, could have confirmed the model. Unfortunately, this could not be realised within the timeframe of the research.

The activity of all the activating esters was determined using HPLC analysis, which is a time consuming technique. Nevertheless, it was the method of choice to distinguish between synthesis and hydrolysis and quantify the products. It would have been more convenient to use a spectrophotometrical assay, which could directly demonstrate synthesis instead of hydrolysis. Such a development would have enabled high throughput screening of many enzymes with our set of esters, and might have resulted in better combinations of enzymes and activating groups than we identified so far.

8.4 Restrictions to large-scale implementation of activating esters

The objective to explore the scope and limitations of the substrate mimetics strategy, which should ultimately have led to a universal approach to couple amino acids to a peptide chain or effect fragment couplings using cheap enzymes, appeared to be more challenging than expected.

First of all, we showed that in some situations it is better to use the term activating esters instead of substrate mimetics, although both strategies have a similar potential for becoming a universal and versatile approach.

Secondly, for every enzyme the search for suitable activating esters has to be carried out separately, as the properties of each enzyme are different. The S/H ratio is enzyme dependent and is moreover varying for each amino acid. An intrinsic problem is that specificity is the main reason to use an enzyme, but at the same time is also its most limiting characteristic. Possibly, the OCam and OTfe esters have the best chances for being universal activating esters, since they are relatively small and will therefore fit in many active sites. Furthermore, they possess the right electronic properties.

In the third place, an essential condition for activating esters becoming part of an industrial process is to develop an effective system to install the activating ester. At this moment all the advantages of enzymatic peptide coupling are lost, because the ester has to be synthesised by chemical methods. In this situation, it would be more efficient to accomplish the peptide bond directly. Without a doubt, the strategy of substrate mimetics and activating esters is only worthwhile, when the ester itself is synthesised in an enzymatic process.

In conclusion, the development of a universal method for chemoenzymatic peptide synthesis has appeared troublesome. However, the synthesis of a particular peptide product using enzyme-specific activation in my view may still be feasible. In that case, the focus is shifted from general difficulties to finding a solution for specific challenges occurring in that particular sequence.

8.5 Outlook

In my opinion, the most promising strategy for developing enzyme-specific activation procedures would consist of a combination of approaches. Several issues have to be tackled, the main two being 1) the still existing hydrolytic capacity of the enzyme, and 2) the recognition of a broad range of amino acid substrates. The problem of undesired enzymatic hydrolysis resulting in low S/H ratios can be solved by genetic engineering. The conversion of the protease subtilisin into subtiligase (S221C and P225A) is a very successful example of such an approach.[3] To prevent secondary hydrolysis of peptide products containing a specific amino acid residue recognised by the enzyme, another approach is required. Ideally, using directed evolution protocols an enzyme is created, which is specific for an 'orthogonal' amino acid, *i.e.* a residue not likely to be incorporated in any peptide chain. The previously discussed spectrophotometrical assay to demonstrate synthesis instead of hydrolysis is essential at this point to enable high throughput screening of the clones. For the resulting reprogrammed enzyme, a specifically activating group should be developed to achieve more universal substrate acceptance. All in all, a challenging task.

8.6 References

[1] (a) L. Edens, P. Dekker, R. van Hoeven, F. Deen, A. de Roos, R. Floris, *J. Agric. Food Chem.* **2005**, *53*, 7950-7957; (b) M. Lopez, L. Edens, *J. Agric. Food Chem.* **2005**, *53*, 7944-7949.

[2] D. V. Besedin, G. N. Rudenskaya, *Russ. J. Bioorg. Chem.* **2003**, *29*, 1-17.

[3] T. Chang, D. Y. Jackson, J. P. Burnier, J. A. Wells, *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 12544-12548.

Summary

This thesis describes a section of the chemoenzymatic peptide synthesis project, which has been performed within the framework of the NWO-IBOS programme. The integration of Biosynthesis and Organic Synthesis is the central approach to synthesise peptides in a sustainable and efficient way. This is important, because peptides make up a growing segment of drugs within the pharmaceutical industry, but the development of improved production methods lags behind.

Peptides are composed of amino acids, which can be coupled both in a chemical and enzymatic way. An overview of these procedures is provided in **Chapter 1**. In addition, special attention is paid to the substrate mimetics strategy, which was taken as a starting point for the research described in this thesis. The substrate mimetics strategy combines the advantages of chemical synthesis (almost unlimited choice of amino acids) with the benefits of enzymatic synthesis (region- and stereospecific, mild reaction conditions). In practice, this is realised by equipping amino acids with an additional universal moiety, the so-called mimetic, resulting in recognition by the enzyme.

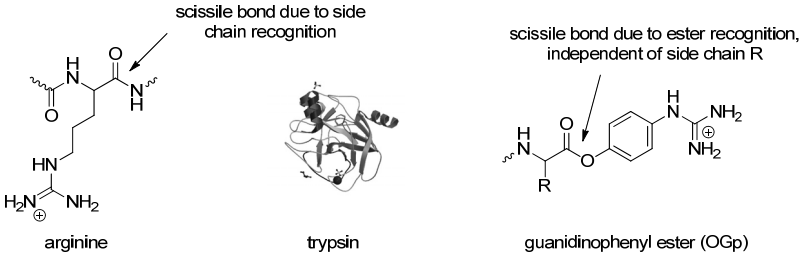
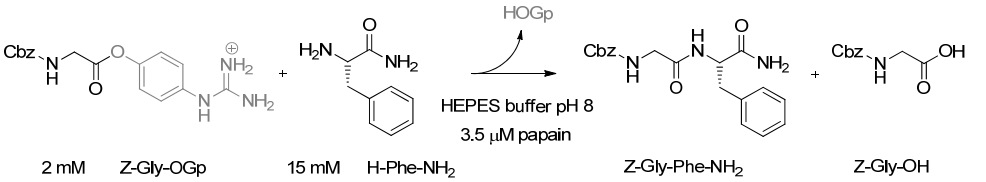


Figure 1 *Principle of the substrate mimetics strategy*

A well-known mimetic from literature is the guanidinophenyl ester (OGp), which, by resembling the amino acid arginine, is specifically recognised by the enzyme trypsin (Figure 1). Until now, OGp has been mainly applied in small-scale academic research. However, to make this concept of substrate mimetics applicable for large-scale industrial processes, a simple and cheap replacement of OGp is required. Preferably, it should be possible to install the mimetic enzymatically in the amino acid of choice.

Our quest for an alternative mimetic commenced in a fairly fundamental way by investigating whether OGp would prove a suitable mimetic for papain, a cysteine protease with a broader specificity than trypsin (**Chapter 2**). A less specific enzyme may allow more variation in the structure of the mimetic without losing activity. As a test substrate Z-Gly-OGp was synthesised and subjected to a chemoenzymatic reaction with papain in the presence of H-Phe-NH_2 as acyl acceptor under aqueous conditions (Scheme 1). The reaction was followed in time and analysed using HPLC.



Scheme 1 *Chemoenzymatic synthesis with papain*

To our delight, the conversion into the corresponding dipeptide was very efficient, which led us to conclude that OGp coupled to Z-Gly-OH is well-recognised by papain. To gain insight into the molecular interactions between enzyme and substrate mimetic, a modelling study was performed, in which also the remaining natural amino acids were involved, as well as a β- and a D-configured amino acid. Surprisingly, this study showed an unexpected binding mode: OGp was predicted to bind papain in a different orientation than the arginine side chain. This suggests that, in this case, the OGp ester is not actually mimicking the natural substrate, but recognition is accomplished in an alternative manner. We named this phenomenon enzyme-specific activation.

Based on these results, we hypothesised that it is not strictly required to have an arginine-like activating ester, provided that sufficient interactions can be made with the enzyme. In **Chapter 3** this hypothesis is investigated, using a set of potentially activating moieties (Figure 2). The variation was mainly brought about by the presence or absence of a positive charge under slightly basic conditions, an aromatic or aliphatic part and the possibility to form hydrogen bonds.

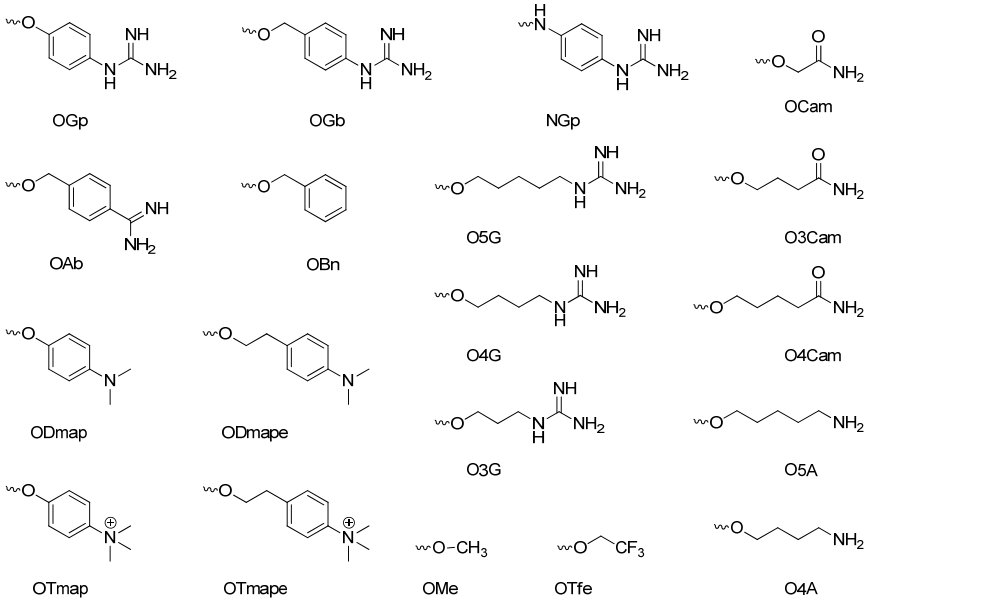


Figure 2 *Structures of potentially activating moieties*

Each of these moieties was coupled to Z-Gly-OH, experimentally evaluated with papain, and subjected to modelling studies, which showed that OBn and ODmap are the most suitable replacements for OGp. Besides Z-Gly-OH, the remaining natural amino acids and a β- and a D-configured amino acid can indeed be used to chemoenzymatically synthesise dipeptides, albeit in varying yields and with an exception for Z-Pro-OH. The applicability of the ODmap and OBn esters was validated in preparative-scale syntheses of two biologically active dipeptides. The computational model demonstrated that besides a proper fit in the active site, the leaving group ability is also important. This intricate relationship was studied further in **Chapter 4** by means of OGp analogues (Figure 3) and the enzyme trypsin, using both experimental and computational approaches.

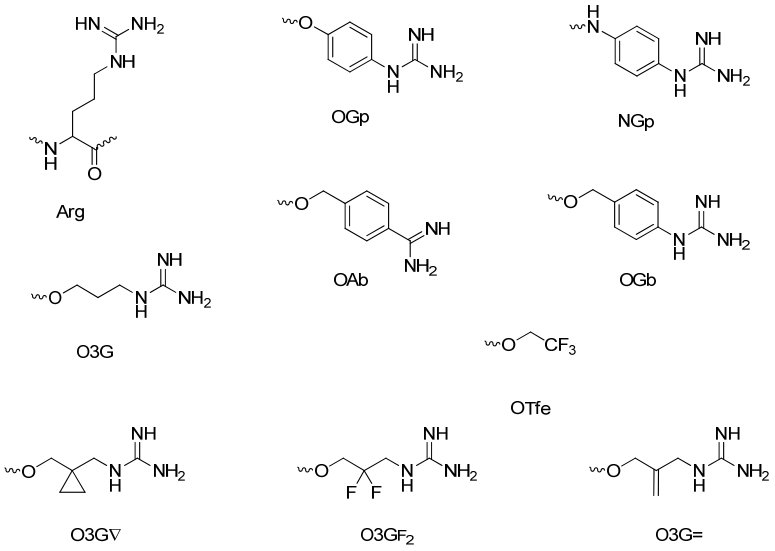


Figure 3 *Structures of OGp analogues*

The experimental evaluation presented us with large differences in reaction times in the conversion of various Z-Gly-analogues into dipeptides in the presence of H-Phe-NH_2 under the influence of trypsin. Based on these observations however, it is virtually impossible to distinguish between the contribution of affinity for the enzyme (a proper fit) and the contribution of the leaving group ability (electronic properties) on the reaction speed. Docking and *ab initio* calculations, two computational techniques, provided insight into the fit of the substrate in the enzyme and the leaving group ability, respectively. Combining all data taught us that the electronic properties of the activating moiety are decisive for enzymatic activity. However, as a precondition, the activating moiety has to fit in the active site, and a nice fit certainly contributes to a higher activity.

Following this study to determine which properties contribute to a successful activating moiety, we investigated the scope of enzyme-specific activation. The previously introduced set of potentially activating moieties (Figure 2) was employed to also find a suitable

activating group for α-chymotrypsin (**Chapter 5**). Furthermore, a comparison was made with the results obtained with papain. It was observed that OBn and ODmap, activating moieties for papain, do not have a specific activating effect on α-chymotrypsin, whereas the reverse is true for OGb. The activating potential of the different moieties are indeed enzyme dependent.

A similar study was carried out with the enzyme Alcalase-CLEA, except that we switched to an organic solvent (**Chapter 6**). Under these conditions the best performing moieties are the two positively charged phenols OGp and OTmap, which is remarkable for an enzyme that prefers large uncharged amino acid residues. Yet, these two moieties happen to be good leaving groups, which is in line with the outcome of chapter 4.

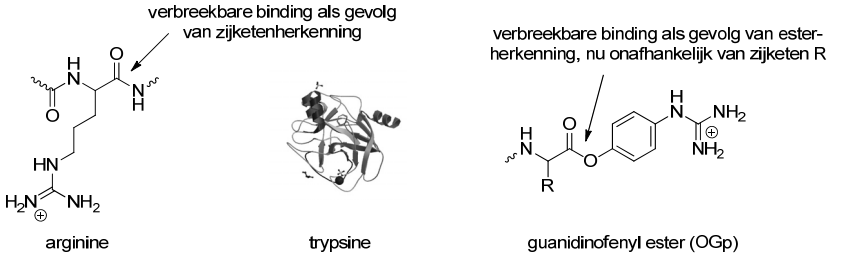
Chapter 7 covers the synthesis of all activating moieties that appear in this thesis.

In **Chapter 8** eventually, my personal view on this piece of research and the results are described. A better fundamental understanding of activating esters may be the most important result produced. Based on this insight I wonder whether an industrially cost-effective process for peptide synthesis derived from substrate mimetics or activating esters is feasible in the near future. Nevertheless, I am convinced that the scope and limitations are extensively mapped out.

Samenvatting

Dit proefschrift beschrijft een onderdeel van het chemoenzymatische peptidesynthese project dat uitgevoerd is binnen het kader van het NWO-IBOS programma. De Integratie van Biosynthese en Organische Synthese staat centraal met als doel peptiden op een duurzame en efficiënte wijze te synthetiseren. Dit is belangrijk omdat peptiden een groeiend segment vormen binnen de farmaceutische industrie, maar de ontwikkeling van verbeterde productiemethoden achterloopt bij de vraag.

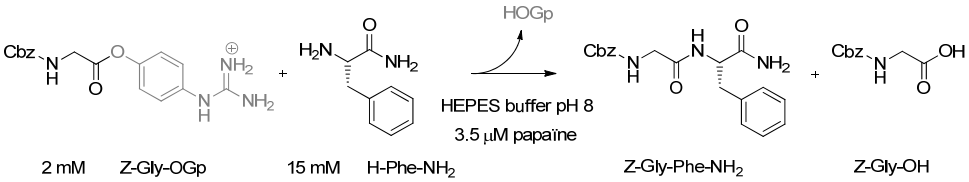
Peptiden zijn opgebouwd uit aminozuren, die zowel chemisch als enzymatisch gekoppeld kunnen worden. In **hoofdstuk 1** wordt hier verder op ingegaan. Daarnaast is er speciale aandacht voor de *substrate mimetics* strategie, die als uitgangspunt heeft gediend voor dit proefschrift. In de *substrate mimetics* strategie worden de voordelen van chemische synthese (nauwelijks gelimiteerde keuze van aminozuren) gecombineerd met de voordelen van enzymatische synthese (regio- en stereospecifiek, milde reactiecondities). Dit wordt bewerkstelligd door aminozuren toe te rusten met een extra universele groep, de zogenaamde *mimetic*, wat herkenning door een enzym tot gevolg heeft.



Figuur 1 *Principe van de substrate mimetics strategie*

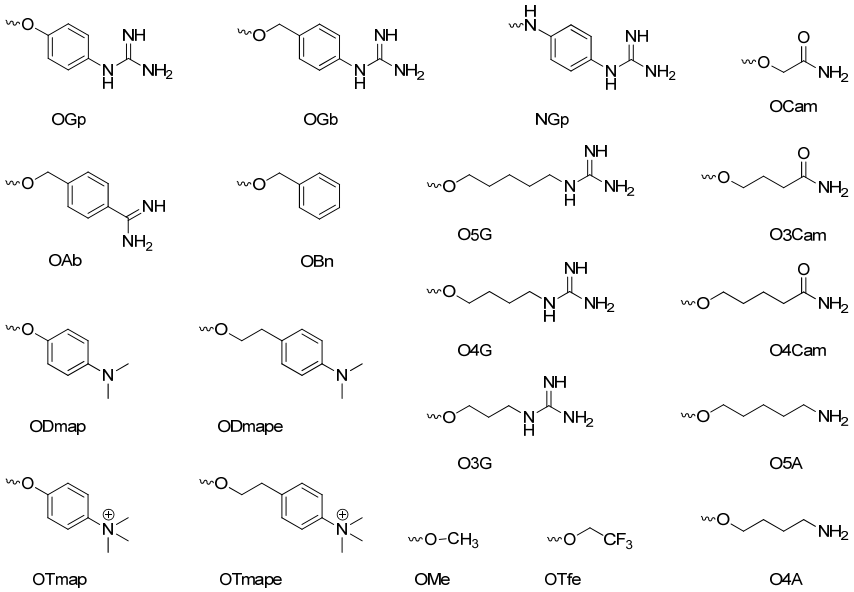
Een bekende *mimetic* uit de literatuur is de guanidinofenyl ester (OGp), die lijkt op het aminozuur arginine dat specifiek herkend wordt door het enzym trypsine (Figuur 1). Tot nu toe is OGp voornamelijk toegepast in academische studies op kleine schaal. Echter, voor een industrieel proces is een versimpelde en goedkope vervanging van OGp nodig, die zo mogelijk zelf ook weer enzymatisch kan worden geïnstalleerd.

Onze zoektocht naar een alternatieve *mimetic* begon redelijk fundamenteel door te onderzoeken of OGp ook een geschikte *mimetic* zou zijn voor papaïne, een cysteïneprotease met een bredere specificiteit dan trypsine (**hoofdstuk 2**). Een enzym dat minder specifiek is laat waarschijnlijk meer variatie toe in de *mimetic* groep, zonder aan activiteit in te boeten. Bij wijze van testsubstraat is Z-Gly-OGp gesynthetiseerd en onderworpen aan een chemoenzymatische reactie met papaïne in de aanwezigheid van H-Phe-NH_2 als acylacceptor in een waterige omgeving (Schema 1). De reactie werd gevolgd in de tijd en geanalyseerd met behulp van HPLC.



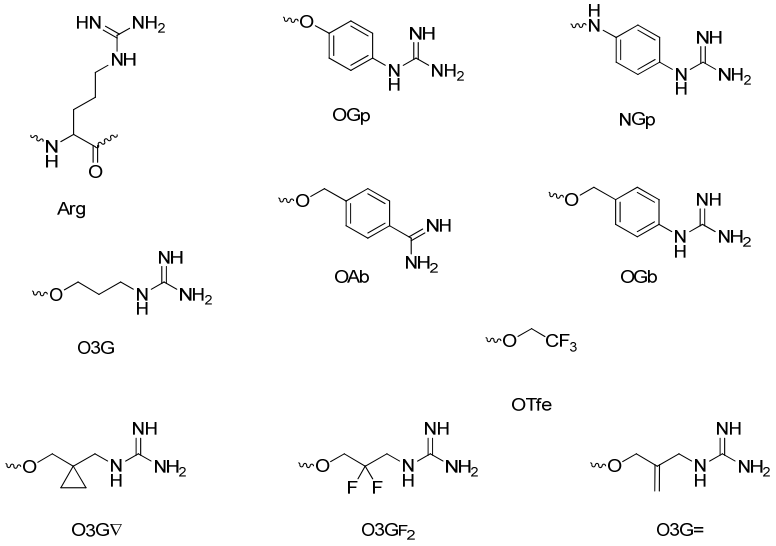
Schema 1 *Chemoenzymatische synthese met papaïne*

Tot onze vreugde verliep de omzetting zeer efficiënt en mag worden geconcludeerd dat OGp gekoppeld aan Z-Gly-OH goed herkend wordt door papaïne. Om een beter beeld te krijgen van de moleculaire interacties tussen enzym en *substrate mimetic* is een modelleerstudie uitgevoerd, waarin ook de overige natuurlijke aminozuren betrokken zijn, evenals een β- en D-aminozuur. Verrassend genoeg resulteerde deze studie in een onverwachte bindingswijze van het substraat: OGp lijkt op een andere manier aan papaïne te binden dan de zijketen van arginine. Dit suggereert dat OGp in dit geval niet optreedt als directe *look-a-like* van arginine, maar op een alternatieve wijze enzymatische herkenning tot stand brengt. Dit fenomeen hebben we enzym-specifieke activering genoemd.
Op basis van het voorgaande, verwachtten we dat een activerende groep voor papaïne niet noodzakelijkerwijs op arginine hoeft te lijken, zolang deze maar voldoende specifieke interacties kan maken met het enzym. In **hoofdstuk 3** wordt deze hypothese getest, uitgaande van een set potentieel activerende groepen (Figuur 2). De variatie zit voornamelijk in de aan- of afwezigheid van een positieve lading onder licht basische condities, een aromatisch of alifatisch gedeelte en de mogelijkheid om waterstofbruggen te vormen.



Figuur 2 *Structuren van potentieel activerende groepen*

Elk van deze groepen is gekoppeld aan Z-Gly-OH, experimenteel getest met papaïne en gemodelleerd, waarna bleek dat OBn en ODmap de beste alternatieven vormen voor OGp.
Naast Z-Gly-OH kunnen ook de overige natuurlijke aminozuren en een β- en D-aminozuur gebruikt worden om chemoenzymatisch dipeptiden mee te synthetiseren, zij het in een variërende opbrengst en met uitzondering van Z-Pro-OH. De toepasbaarheid van ODmap en OBn is geïllustreerd met de synthese van twee biologisch actieve dipeptiden op preparatieve schaal. Uit het computationele werk kwam naar voren dat het belangrijk is dat de activerende groep goed past in het enzym, maar ook dat de intrinsieke reactiviteit voldoende groot is.
Dit laatste aspect is uitgebreider onderzocht in **hoofdstuk 4** met behulp van OGp analoga (Figuur 3) en het enzym trypsine, gebruikmakend van zowel experimentele als computationele methoden.



Figuur 3 *Structuren van OGp analoga*

De experimentele evaluatie liet grote verschillen zien in de snelheid waarmee de verschillende Z-Gly-analoga onder invloed van trypsine en in aanwezigheid van H-Phe-NH_2 werden omgezet in dipeptiden. Het is echter vrijwel onmogelijk om op basis hiervan onderscheid te maken tussen de bijdrage van affiniteit voor het enzym (de pasvorm) en de bijdrage van het vertrekkende groep karakter (elektronische eigenschappen) op de reactiesnelheid. *Docking* en *ab initio* berekeningen, twee computationele technieken, geven meer inzicht in respectievelijk de pasvorm van het substraat in het enzym en het vertrekkende groep karakter. Combinatie van alle data leert dat de elektronische eigenschappen van de activerende groep een grotere invloed hebben op de enzymatische activiteit, mits de activerende groep past in het enzym. Wel geldt dat een betere pasvorm resulteert in een hogere enzymatische activiteit.

Na deze studie om te bepalen welke eigenschappen tot een succesvolle activerende groep leiden, is vervolgens gekeken naar de bredere toepasbaarheid van enzym-specifieke activering. De eerder gebruikte set potentieel activerende groepen (Figuur 2) is aangewend om ook voor α-chymotrypsine een geschikte groep te vinden (**hoofdstuk 5**). Bovendien is een vergelijking gemaakt met de resultaten met papaïne. Opvallend is dat OBn en ODmap, activerende groepen voor papaïne, geen specifiek activerend effect hebben op α-chymotrypsine, terwijl het omgekeerde waar is voor OGb. De activerende eigenschappen van de groepen zijn dus daadwerkelijk enzym-afhankelijk.
Een vergelijkbaar onderzoek heeft plaatsgevonden voor het enzym Alcalase-CLEA, met een additionele grote verandering, namelijk de overstap naar een organisch oplosmiddel (**hoofdstuk 6**). In dit geval zijn de best presterende groepen de twee positief geladen fenolen OGp en OTmap, wat opmerkelijk is voor een enzym met een voorkeur voor grote ongeladen hydrofobe residuen. Het zijn echter wel twee goed vertrekkende groepen, wat weer in lijn is met de uitkomsten van hoofdstuk 4.
Hoofdstuk 7 beslaat de synthese van alle activerende groepen die voorkomen in dit proefschrift.
In **hoofdstuk 8** ten slotte komen mijn persoonlijke visie op dit onderzoek en de uitkomsten ervan aan bod. Een beter fundamenteel inzicht in de werking van *substrate mimetics* is misschien wel het belangrijkste wat dit proefschrift heeft voortgebracht. Op basis daarvan vraag ik me af of het op korte termijn tot een industrieel rendabel proces voor peptidesynthese gebaseerd op *substrate mimetics* of activerende esters zal komen. Ik ben er echter van overtuigd dat de aandachtspunten en mogelijke problemen uitgebreid in kaart zijn gebracht.

Dankwoord

Graag wil ik je bedanken door mijn favoriete gedicht met je te delen:

Ik ben lekker stout

Ik wil niet meer, ik wil niet meer!
Ik wil geen handjes geven!
Ik wil niet zeggen elke keer:
Jawel mevrouw, jawel meneer…
nee, nooit meer van m'n leven!
Ik hou m'n handen op m'n rug
en ik zeg lekker niks terug!

Ik wil geen vieze havermout,
ik wil geen tandjes poetsen!
'k Wil lekker knoeien met het zout,
ik wil niet aardig zijn maar stout
en van de leuning roetsen
en schipbreuk spelen in de teil
en ik wil spugen op het zeil!

En heel hard stampen in een plas
en dan m'n tong uitsteken
en morsen op m'n nieuwe jas
en ik wil overmorgen pas
weer met twee woorden spreken!
En ik wil alles wat niet mag,
De hele dag, de hele dag!

En ik wil op de kanapee
met hele vuile schoenen
en ik wil aldoor gillen: Nee!
En ik wil met de melkboer mee
en dan het paardje zoenen.
En dat is alles wat ik wil
en als ze kwaad zijn, zeg ik: Bil!

Uit: Annie M.G. Schmidt, *Ik ben lekker stout,* 1955, Em. Querido's Uitgeverij Amsterdam

Acknowledgement

I would like to thank you by sharing my favourite poem with you:

Nice and Naughty

I've had enough! I'm not a lamb!
I don't want to say hello!
I don't want to say, 'Yes, sir,' – 'Yes, ma'am,'
or hear about how big I am...
as if I didn't know.
I'll stay out in the yard all day
until I'm sure they've gone away!

I don't want to eat that mushy rice,
I don't want to brush my hair!
I don't want to hear their good advice.
I want to be naughty, not nice,
and lean back on my chair,
and play pirates in the kitchen sink,
and fi nger-paint with Indian ink!

And when it rains I'll go and dance
in puddles up to my knees.
I'll poke my tongue out at my aunts,
'cause when I'm old I'll have a chance
to say thank you and please.
And I'll do everything that's wrong,
the whole day long, the whole day long!

I want to jump on the settee
and cover it with grime.
I want to scream hysterically,
and take the dog to bed with me...
but I'll say when it's time.
That's all the things I plan to do.
If they don't like it, I'll say, 'Poo!'

Translated by: David Colmer

Curriculum Vitae

Roseri de Beer is geboren op 6 juni 1983 te Tilburg in Nederland, maar bracht de eerste jaren van haar leven door in Oosterhout (NB). Vervolgens verhuisde de familie naar Gorinchem, waar zij naar het Gymnasium Camphusianum ging (1995-2001, cum laude).
Op achttienjarige leeftijd vertrok zij naar Wageningen om Moleculaire Wetenschappen (2001-2004, BSc cum laude) te gaan studeren aan de WUR. De meeste vrije tijd besteedde zij bij studentenvereniging SSR-W, waar ze activiteiten organiseerde en deelnam aan verschillende commissies. In '04-'05 onderbrak Roseri haar studie voor een fulltime bestuursjaar bij SSR-W als Commissaris Mensa. Gedurende haar master (2005-2007, MSc cum laude) besloot ze om naar het buitenland te gaan voor een tweede afstudeervak in Madrid en een stage in Basel.
Na haar afstuderen begon zij aan een promotieonderzoek in de Synthetische Organische Chemie groep van prof. dr. Floris Rutjes aan de Radboud Universiteit te Nijmegen (2008-2012). De behaalde resultaten staan beschreven in dit proefschrift.
Momenteel is zijn aangesteld als tijdelijk onderzoeker bij het klinisch-chemisch en hematologisch laboratorium in Ziekenhuis Gelderse Vallei te Ede. In de nabije toekomst hoopt zij een opleidingsplaats tot klinisch chemicus te vinden.

Curriculum Vitae

Roseri de Beer was born on the 6th of June 1983 in Tilburg, the Netherlands, but lived in Oosterhout (NB) the first couple of years. Subsequently, the family moved to Gorinchem where she attended Gymnasium Camphusianum (1995-2001, cum laude).
On the age of eighteen she moved to Wageningen to study Molecular Sciences (2001-2004, BSc cum laude) at the WUR. She spent most of her free time at student association SSR-W, organising activities and participating in several committees. In '04-'05, Roseri interrupted her studies to be a fulltime board member of SSR-W, being responsible for the Mensa. During the master's degree course (2005-2007, MSc cum laude) she decided to go abroad for a second thesis in Madrid and an internship in Basel.
After graduation, she started with a PhD project in the Synthetic Organic Chemistry group of prof. dr. Floris Rutjes at the Radboud University in Nijmegen (2008-2012). The results obtained are described in this thesis.
Currently, she is appointed as temporal researcher at the clinical chemistry laboratory in hospital Gelderse Vallei in Ede. In the near future she hopes to find a position in which she is trained to become a clinical chemist.